Common Media Application Format (CMAF)

# Introduction

There has been an explosive growth of Internet delivered video traffic, a growth predicted to continue for years to come. According to the 2017 Cisco Visual Networking Index, by 2021 it would take you more than 5 million years to watch the amount of video that will cross global IP networks each month, IP video traffic will be 82% of all consumer internet traffic, 63% of total IP traffic will be going to wireless and mobile devices and the number of IP connected devices will be three times the global population. This volume of IP video traffic, when combined with an extraordinary number of targeted IP connected devices presents the video industry with a profound interoperability problem.

MPEG has published several specifications to address this interoperability challenge, including Dynamic Adaptive Streaming over HTTP and the Common Encryption specifications. The Common Media Application Format (CMAF) is the extensible encoding standard which completes this picture, codifying and enhancing what has become best practice for IP video distribution, enabling interoperability between content and devices by specifying consistent media packaging and encryption.

CMAF is optimized for efficient large scale protected video applications, such as streaming televised events, where CMAF Segments can be delivered once to edge servers in content delivery networks, then accessed from cache by thousands of streaming video players without requiring additional network backbone traffic or transmission delay.

The CMAF features which make this possible:

* **CMAF uses Common Encryption** which enables the same encrypted presentation to be protected by different DRM systems built into different devices. Common Encryption works with standard HTML5 APIs, enhancing application interoperability.
* CMAFdefines interoperable **CMAF Media Profiles** that are identified by registered brands. These Media Profiles specify the codec and encoding constraints necessary for decoding and rendering interoperability, as well as requirements to ensure seamless track switching necessary for dynamic adaptive streaming.
* **CMAF is extensible**. New Media Profiles can be derived by reference to the generic CMAF Fragment, Track, and Switching Set formats specified in the core standard, in combination with Media Profile specific codec and encoding constraints identified by a registered ISOBMFF brand.
* **CMAF Fragments** enable the encoding and decoding of CMAF Media Profiles independent from the delivery method.
* CMAF Fragments can be packaged in larger or smaller **CMAF Addressable Media Objects** for delivery purposes. Larger **CMAF Segments** can be used to optimize network efficiency where added delay is acceptable, or smaller **CMAF Chunks** can be used to stream the media samples in a Fragment before the entire Fragment has been encoded to reduce presentation delay in live streaming.
* **CMAF Switching Set** constraints enable the use of a single-track buffer and decoder for each Switching Set so that CMAF content can be seamlessly switched on most deployed devices and browsers.
* **CMAF Presentations** can be played on billions of Internet devices, using code in a Web page, a player included in a Web browser, or a player embedded in a device.

# A Deeper Dive into CMAF Technology

CMAF Media Objects and Tracks are derived from ISO Base Media File Format (ISOBMFF) using movie fragments with movie fragment relative byte range addressing and a decode timestamp to sequence and synchronize each movie fragment in a CMAF Presentation. CMAF Media Objects enables splicing and parsing of CMAF Fragments independent of its storage and delivery, enabling live streaming of non-persistent CMAF Fragments as well as adaptive sequencing of CMAF Fragments from different CMAF Tracks to adapt to variations in network throughput.

CMAF specifies each CMAF Fragment to contain the metadata necessary to decrypt, decode, and display it when processed according to the constraints of CMAF Tracks, Switching Sets, Selection Sets, and Presentations. The media samples in CMAF Fragments are constrained to be independently encodable and decodable, and to conform to the codec and encoding parameter limits of a CMAF Media Profile.

CMAF encoding and decoding of CMAF Fragments is independent of delivery objects and protocols, allowing storage and delivery of CMAF Fragments packaged as complete files, streams of independently decodable segments, or live streams that package Fragments in small chunks for low delivery delay. CMAF Chunk encoding allows CMAF Segment requests to be progressively delivered as each CMAF Chunk completes encoding and packaging, rather than waiting to encode and package all the samples in the Fragment. Later requests or unoptimized systems will receive a single Segment in response to a request, and that Segment containing may contain a CMAF Fragment containing that contains a sequence of CMAF Chunks.

CMAF enables each player to select and combine Tracks during playback by storing each media component in a separate CMAF Track, and specifying how CMAF Tracks are to be start aligned and synchronized. Each player can select and download different media components conforming to CMAF Selection Sets, such as alternative languages, codecs, bitrates, and video resolutions; optimized for different users, devices, and network conditions.

If each combination of media components had to be early bound in a multiplexed prior to delivery file (i.e. each file containing one combination of audio, video, and subtitles to minimize the bandwidth required), that could result in hundreds of multitrack files, which would greatly increase network storage and caching and bandwidth requirements. The number of files would be the product of the number of languages, content types (dialog, description, commentary, etc.), audio codecs, codec profiles and levels, sample formats, audio channel configurations, video resolutions, video characteristics (standard or high dynamic range, wide color gamut, etc.), etc.; times the number of video Media Profiles, bitrates, resolutions, frame rates, encryption modes, etc. Multiplexing of elementary streams in other media formats results in redundant packaging, storage, and transmission of duplicate elementary streams to edge servers, and greatly reduces cache efficiency by increasing the number of alternative segments.

Single track CMAF Segments and late binding allow the same audio CMAF Segment, for example, to be used by players that are playing a variety of video CMAF Fragments with different bitrates, resolutions, and codecs. Since each CMAF Fragment is stored and cached only once, network bandwidth, storage, and cache are more efficiently utilized.

# CMAF Media Objects and Terminology

CMAF specifies the following **CMAF Media Objects** and system model:

23000-19_ed1fig1

**CMAF Tracks** contain encoded media samples, including audio, video, and subtitles, stored in a CMAF specified container and brand derived from the ISO Base Media File Format (ISOBMFF). CMAF Tracks may exist temporarily while streaming, or may be stored.

**Media Samples** may optionally be protected by **MPEG Common Encryption**.

**CMAF Tracks** consists of a **CMAF Header** followed by **CMAF Fragments** that contain the media samples.

**CMAF Fragments** are independently decodable and decryptable, in combination with their associated CMAF Header.

**CMAF Switching Sets** contain alternative CMAF Tracks that can be switched and spliced at CMAF Fragment boundaries to adaptively stream the same content at different bitrates and resolutions.

**CMAF Selections Sets** contain alternative CMAF Switching Sets that may include alternative content, e.g. different languages or camera angles, or alternative encodings, e.g. different codecs.

**CMAF Addressable Media Objects** are specified for CMAF Track storage and delivery, including **CMAF Chunks** (smaller than a CMAF Fragment), **CMAF Segments** (one or more CMAF Fragments), and **CMAF Track Files** (a complete CMAF Track in one ISOBMFF file).

The **CMAF Hypothetical Reference Model** defines how CMAF Tracks can be delivered, combined, and synchronized in **CMAF Presentations**; but the model allows the use of any compatible implementation, including broadcast and MPEG DASH adaptive streaming.