|  |  |
| --- | --- |
|  |  |

Text

Description automatically generated ISO/IEC JTC 1/SC 29/WG 03 N1605

**ISO/IEC JTC 1/SC 29/WG 03**

**MPEG Systems   
Convenorship: KATS (Korea, Republic of)**

**Document type:** Output Document

**Title: Exploration on media authenticity and provenance indication with the MPEG Systems technologies**

**Status:** Approved

**Date of document:** 2025-10-07

**Source:** ISO/IEC JTC 1/SC 29/WG 03

**No. of pages:** 21 (with cover page)

**Email of Convenor:** young.L@samsung.com

**Committee URL:** <https://isotc.iso.org/livelink/livelink/open/jtc1sc29wg3>

**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 03 MPEG SYSTEMS**

**ISO/IEC JTC 1/SC 29/WG 03 N** **1631**

**Oct 2025, Geneva, CH**

|  |  |
| --- | --- |
| **Source** | **MPEG Systems working group** |
| **Status** | **Output** |
| **Title** | **Exploration on media authenticity and provenance indication with the MPEG Systems technologies** |
| **Serial Number** | **25605** |

**Table of Contents**

[1 Introduction 3](#_Toc211603237)

[2 Background 3](#_Toc211603238)

[3 System level requirements on technologies for multimedia authentication 3](#_Toc211603239)

[4 Definitions 4](#_Toc211603240)

[5 Scope of work 5](#_Toc211603241)

[6 Producers and receivers 5](#_Toc211603242)

[7 Use cases 5](#_Toc211603243)

[7.1 Misinformation and disinformation 6](#_Toc211603244)

[7.1.1 Deepfake detection by authenticity verification 6](#_Toc211603245)

[7.1.2 Manipulated media 6](#_Toc211603246)

[7.1.3 Virtual meeting authentication 6](#_Toc211603247)

[7.1.4 Impersonation using video and audio 6](#_Toc211603248)

[7.1.5 Fake news with synthesized audio 6](#_Toc211603249)

[7.2 Media forgery 6](#_Toc211603250)

[7.2.1 Insurance fraud 6](#_Toc211603251)

[7.2.2 Surveillance footage integrity 6](#_Toc211603252)

[7.2.3 Audio recordings in investigations 6](#_Toc211603253)

[7.3 Media creation 7](#_Toc211603254)

[7.3.1 Authenticity & integrity for media 7](#_Toc211603255)

[7.3.2 AI-generated content (AIGC) 7](#_Toc211603256)

[7.4 Media modification 7](#_Toc211603257)

[7.4.1 Video and audio editing software 7](#_Toc211603258)

[7.4.2 Media adaptation during transport 7](#_Toc211603259)

[7.4.3 Content splicing and ad-insertion 7](#_Toc211603260)

[7.4.4 Video and audio transcoding for streaming 7](#_Toc211603261)

[7.4.5 Movie and audio preservation 7](#_Toc211603262)

[7.5 Media augmentation and extension 8](#_Toc211603263)

[7.5.1 Media composition 8](#_Toc211603264)

[7.5.2 Identifying augmented reality (AR) and virtual reality (VR) components 8](#_Toc211603265)

[7.6 General characteristics of the use cases 8](#_Toc211603266)

[8 System level requirements 8](#_Toc211603267)

[8.1 General aspects 8](#_Toc211603268)

[8.2 System level specific requirements 9](#_Toc211603269)

[9 Gap Analysis 10](#_Toc211603270)

[9.1.1 m73494 Analysis of systems media authenticity requirements based on MPEG-7 and MPEG-21 10](#_Toc211603271)

[9.1.2 m73485 Gap analysis of ISOBMFF and DASH to support media authentication requirements 10](#_Toc211603272)

[9.1.3 m73486 Gap analysis of C2PA to support media authentication requirements in ISOBMFF and DASH 10](#_Toc211603273)

[9.2 m73486 Gap analysis of C2PA to support media authentication requirements in ISOBMFF and DASH 11](#_Toc211603274)

[10 Workflows (based on m74553 and m7440) 11](#_Toc211603275)

[10.1 Streaming Workflow (m74553) 11](#_Toc211603276)

[10.2 File-based workflow (m74553) 12](#_Toc211603277)

[10.3 Cross authentication in the file-based delivery workflow (m74440) 14](#_Toc211603278)

[10.4 Content properties (m74553) 15](#_Toc211603279)

[10.5 Authentication properties (m74553) 15](#_Toc211603280)

[10.6 Interfaces (m74553) 15](#_Toc211603281)

[10.7 Authentication functions (m74553) 16](#_Toc211603282)

[10.7.1 Authentication author 16](#_Toc211603283)

[10.7.2 Authentication verifier 16](#_Toc211603284)

[10.8 Interfaces’ features (m74553) 16](#_Toc211603285)

[10.8.1 AIF-1a 16](#_Toc211603286)

[10.8.2 AIF-1b 16](#_Toc211603287)

[10.8.3 AIF-1c 17](#_Toc211603288)

[10.8.4 AIF-2a 17](#_Toc211603289)

[10.8.5 AIF-2b 17](#_Toc211603290)

[10.8.6 AIF-2c 17](#_Toc211603291)

[10.8.7 AIF-3c 18](#_Toc211603292)

[11 Potential work items (m74553) 18](#_Toc211603293)

[11.1 ISOBMFF 18](#_Toc211603294)

[11.1.1 System-level authentication information in tracks/track fragments 18](#_Toc211603295)

[11.2 Elementary authentication information in tracks 18](#_Toc211603296)

[11.3 DASH 18](#_Toc211603297)

[11.3.1 Signaling at the manifest 18](#_Toc211603298)

[12 Proposed candidate solutions 19](#_Toc211603299)

[12.1 Media cross authentication signaling in ISOBMFF (m74440) 19](#_Toc211603300)

[12.2 Signaling in MPEG-2 TS (m74512) 19](#_Toc211603301)

[13 Discussions at MPEG152 19](#_Toc211603302)

[14 Recommendations 20](#_Toc211603303)

[15 References 20](#_Toc211603304)

# Introduction

This document defines the scope of a WG03 exploration on media authentication and AI-generated/altered content marking in the MPEG Systems standards.

Media authentication, i.e. verification of the truthfulness of media content, through various means has been the subject of discussion, especially in the light of AI generated and altered content. Various news incidences involving the deepfakes, misinformation and disinformation about politics and elections, and public figures, and even in other popular subjects caused the legal authorities in various countries looking into legal aspects of media authentication and the need to require the means of authentication of a publish content for some applications and distribution platforms. While those activities are moving forward and their impact will be anticipated, the technical means enabling media authentication are highly desired to implement those legal requirements as well as the market needs.

# Background

The need for media authentication and AI generated/altered content marking has been raising recently. Legislative works have been started in several jurisdictions that attempt to provide requirements for such solutions.

The recent standard activities such as C2PA and JPEG TRUST address some of the needed functionalities for certain media authentication use cases. However, the ranges of use cases and application are vast. The extend of the use cases and application, as well as various delivery schemes are outline in WG02-NXXXX.

In the technical field, the standard organizations and consortia have been recently investigated or started development of standards around media authentication. In MPEG, WG3 with the exploration on marking AI generated/altered media ([MDS24619\_WG03\_N01411](https://dms.mpeg.expert/doc_end_user/documents/148_Kemer/wg11/MDS24619_WG03_N01411.zip)), and JVET/WG5 with digital signed SEI messages have been working on this topic (JVET-AK2006). Recently, WG02 completed the requirements work on the media authentication (WG02\_N0427), with the goal of creating an umbrella media authentication requirement document for the entire MPEG community. This document defines the main use cases and the technical requirements. Futhermore, contribution m71392 proposed a subset of the WG02\_N00413 requirements for the MPEG system standards and contribution m71125 proposed a method for the use of C2PA with MPEG DASH to address some of the requirements.

# System level requirements on technologies for multimedia authentication

Multimedia content, including text, images, audio, and video, is integral to modern life. Its reliable exchange is supported globally by ISO/IEC JTC1/SC 29 standards. However, its trustworthy usability is increasingly challenged by sophisticated forgeries and manipulations like AI-driven deepfakes. Since such techniques blur the lines between reality and fabrication, they raise significant cybersecurity and trust issues within digital ecosystems.

It is observed that for some of the ISO/IEC JTC1/SC 29 standards like MPEG-D USAC or VVC, no mechanisms for trustworthy verification exist which, on the one hand, can be realized in an independent and self-contained way within these standards but which, on the other hand, can be combined with each other and with other layers of the MPEG-ecosystem. While the first aspect is crucial for applications like independent streaming, the second aspect is important to prevent forgeries via fake combinations of individually verified media assets.

In parallel, the rapid evolution of artificial intelligence (AI) technologies—particularly in large language models, image and video generation, and multimodal systems—has significantly transformed industries like entertainment and healthcare by enhancing productivity and creativity. However, this progress has also led to ethical concerns regarding the proliferation of deceptive content. The growing prevalence of AI-generated content underscores the urgent need for novel verification technologies to maintain media authenticity and prevent malicious usage, restoring trust in digital sources.

This document captures the subset of requirements defined in WG02\_N042 as the system level requirements for the media authentication.

# Definitions

This section provides the definition of terms used in this document.

|  |  |
| --- | --- |
| **Media types** | Types of media including 2-D and multi-component video and visual (such as graphics and animation), audio, closed caption, and haptics. |
| **Media asset** | A media content that may consist of one or more media types, and one or more tracks of each type, which together constitute a media presentation. |
| **Media asset metadata** | The portion of a media asset that provides information about the media asset or its content that is not directly used for rendering the media presentation, such as location, creator, annotations or IPR information [2]. |
| **Media authentication** | Process or action of [proving](https://www.google.com/search?sca_esv=71781198cf3ca03a&rlz=1C1CHBF_enUS1016US1016&sxsrf=ADLYWILlsILEvg3RBGHmBlNCn1xM3mL01w:1729371772807&q=proving&si=ACC90nwUEXg6u2vxy-araGkF9MAxeRxk9DoJIfEkC9933Hcmp8WvZHiAjZdwrzjI8qcF6Z7q-dVFsTO4mdGea_Jcsf7a5MJRWqPjVS3DsvE5G2v28qtPB5g%3D&expnd=1&sa=X&ved=2ahUKEwiLqM_xq5uJAxXvEkQIHQiWJp4QyecJegQIIBAO) or showing a media asset being true, genuine, or valid. |
| **Media authenticity** | The truthfulness and reliability of media content, ensuring that it accurately represents the claimed information or events. |
| **Media integrity** | The preservation of media content in its original state without any unauthorized modifications or corruption during storage or transmission. |
| **Producer** | An entity, person or organization that produces, edits, revises or manipulates a media asset. |
| **Receiver** | An entity, person or organization that receives, retrieves, or streams the content for consumption as the end user or for the purpose of producing another content. |
| **Deepfake** | A media (typically of a person or persons) which has been [digitally](https://www.google.com/search?sca_esv=71781198cf3ca03a&rlz=1C1CHBF_enUS1016US1016&sxsrf=ADLYWIK2o7dlmbtt2RDTfiSeIrS4_NnG_A:1729373489161&q=digitally&si=ACC90nxMSPeZfdJJjQgDsdZJuFuJvXSi0C4w3Vlu-HaJn-nOAP71qiZLqzap8O_x_dvhi0y7XKx3XxUnCS03SqKntBE2CzJtmdTWg6CZY486wUITjcpAeg4%3D&expnd=1&sa=X&ved=2ahUKEwiUn4WkspuJAxVuE0QIHfiyHW4QyecJegQINxAN) altered, typically to be used [maliciously](https://www.google.com/search?sca_esv=71781198cf3ca03a&rlz=1C1CHBF_enUS1016US1016&sxsrf=ADLYWIK2o7dlmbtt2RDTfiSeIrS4_NnG_A:1729373489161&q=maliciously&si=ACC90nwZrNcJVJVL0KSmGGq5Ka2YzdAOXs1-nVE-OfCJgXAgbaBnD8mDsOeQoymcOjTjynmXCPjz9Id67ji7EUdpaHX8GVqOgYBnBCf6vsH1SvR7P4DfqQg%3D&expnd=1&sa=X&ved=2ahUKEwiUn4WkspuJAxVuE0QIHfiyHW4QyecJegQINxAO) or spread false information. |
| **Misinformation** | Information that is based on reality, used to inflict harm on a person, social group, organization or country [1]. |
| **Disinformation** | Information that is false and deliberately created to harm a person, social group, organization or country [1]. |
| **Media forgery** | The intentional manipulation, alteration, or creation of digital media—such as images, videos, or audio—with the purpose of misleading or deceiving viewers about the authenticity or truthfulness of the content. |
| **Trust verification** | The process of verifying that the media is produced by a trusted producer. |

# Scope of work

The scope of this work is the development and/or enhancement of the specification(s) for the inclusion of necessary metadata in the system layer streams, for enabling the authentication of the corresponding media assets. During the development of such standards, MPEG Systems may leverage the existing specifications, when they satisfy parts of the requirements, and develop new metadata formats and procedures to fulfill other requirements for addressing the use cases defined in this document.

# Producers and receivers

Before discussing the use cases, it is important to note to categorize the producers and receivers of the content that is the subject of media authentication.

The producers of the content can be categorized into two major groups:

1. Professional producers (PP): The professional organizations and entities that produce the content, such as news agencies and TV networks.
2. General producers (GP): Individual users that capture or produce content, for instance the social network subscribers who post media content on their account on any social network.

Similarly, the receivers of the media can be grouped into two main categories:

1. General Receiver (GR): A public user usually consumes the media content and is interested in verifying the authenticity of the content. The main task is the authentication and the alternation and editing details of the content is usually not a concern for such user.
2. Professional Receiver (PR): A professional user, such as a journalist, news agency or authority usually has an interest in the origin of the content and the various steps of alternation, editing and manipulation the content went through, and the sources who perform such alternation and editing on the content.

# Use cases

One of the key objectives of this document is to identify use cases that fall under the scope of multimedia authentication and to identify needs for standardization.

The following use cases have been identified as relevant.

## Misinformation and disinformation

### Deepfake detection by authenticity verification

The transfer operation requires confirmation of the identity of the content provider (e.g. in order to determine whether, when a famous person is making questionable claims, the media is genuine and not a deepfake). Generated by PP/GP and used by GR and PR users.

### Manipulated media

A GR user wants to confirm whether media, generated by a PP, has been altered when published (e.g., to confirm the authenticity of a news video that shows an important event or to verify whether an audio attached to a video of some individual actually corresponds to that video and not to an older video of the same individual).

### Virtual meeting authentication

In business or legal virtual meetings, participants want to ensure that none of the voices have been altered or synthesized to impersonate key decision-makers.

### Impersonation using video and audio

Public figures, celebrities, or business executives are often targeted by fake videos and voiceovers to damage their reputations or impersonate them for malicious purposes. Identifying forged videos or synthesized voices is critical for protecting personal and professional identities. (e.g., A celebrity’s voice is faked in an online interview to make controversial statements. The celebrity needs tools to verify the authenticity of the media to refute the forgery.)

### Fake news with synthesized audio

In media reports, audio can be synthetically generated to fake statements from public figures. Verification of audio source and authenticity is essential to prevent the spread of fake news. (e.g., An audio clip allegedly featuring a famous actor making offensive comments circulates online. Media outlets need to verify the audio’s origin to confirm whether it was artificially generated.)

## Media forgery

### Insurance fraud

An insurance company (PR) wants to confirm the authenticity of some media during a claim process (e.g., a video of a car accident), for the content produced by a GP.

### Surveillance footage integrity

Security teams and law enforcement (PR) must validate that surveillance camera footage (PP/GP) has not been tampered with when used in investigations of criminal activities.

### Audio recordings in investigations

Investigators (PR) want to confirm that audio recordings (PP/GP) used in criminal cases (e.g., wiretaps or interviews) have not been edited or distorted to manipulate the context of the conversation.

## Media creation

### Authenticity & integrity for media

Media organizations (PR) want to confirm the authenticity of the received media assets from PP/GPs.

### AI-generated content (AIGC)

With the rise of AI-generated videos and synthesized audio, receivers (GR) and professionals (PR) alike need to know whether media was created or modified using AI software (PP/GP). Identifying AI-generated content is crucial for transparency. (e.g., Viewers want to verify whether a popular video clip was created using AI or is a genuine, human-created piece of content.)

## Media modification

### Video and audio editing software

The extent of editing on video and audio content should be made transparent for the receiver (PR/GR). For example, professional editing tools (PP) used to enhance or alter content should label post-processing details to indicate authenticity. (e.g., A film editor needs to mark the sections of a movie trailer that have been color-corrected and where the audio has been pitch-adjusted for effect.)

### Media adaptation during transport

The throughput of the network might change over time and some techniques may be applied to adapt it to the network conditions while still being able to authenticate the media (e.g., for scalable media codecs some layers might be dropped, or a media stream could be transcoded.) Applicable to both GR and PR.

### Content splicing and ad-insertion

The media might be spliced (e.g. by inserting the advertisement in the main content or added graphics and animation.) The various parts of the original media need to be verifiable for authenticity. Applicable to both GR and PR.

### Video and audio transcoding for streaming

Audio and video transcoding during streaming processes must retain authenticity verification, ensuring that the content hasn’t been altered beyond codec adjustments. (e.g., An online streaming platform needs to verify that the transcoding of audio tracks to lower bitrates for mobile devices has not altered the content’s original integrity.)

### Movie and audio preservation

Compression, transcoding, and other preservation processes applied to classic films and audio recordings must be documented to maintain transparency. This ensures that the archived content can be trusted as faithful to the original. (e.g., A film archive wants to keep track of the transcoding steps applied to preserve an old movie for modern viewing formats, while preserving its original quality.)

## Media augmentation and extension

### Media composition

Judges and police (PR) want to know whether media evidence (GP and/or PP) is synthesized when adjudicating cases (e.g., when an edited video is composed of several ones).

### Identifying augmented reality (AR) and virtual reality (VR) components

In AR and VR content, both video and audio elements may be augmented or generated in real time. Ensuring that these augmented components are clearly distinguished from real-world captures is important for maintaining authenticity.

## General characteristics of the use cases

As can be seen from the described use cases, media authentication typically invokes several media components.

In a typical use case, a content provider wants to distribute a compressed version of a multimedia asset that consists of a combination of one or more video sequences, multiple audio streams (in different languages) and multiple caption streams for the respective audio streams. The content is generated and distributed on the fly, for example in a live streaming scenario. Moreover, users across the world may want to consume different versions of the content. For example, some users may want to consume an individual audio stream or the video stream only or they may want to consume specific versions of such media content depending on their available bandwidth, as, for instance, the video at a particular resolution or audio at a particular language or bitrate. As such, joint authentication should be possible for such media combinations.

Furthermore, as discussed in the use cases above, only parts or substreams of the original video or audio streams might be received depending on the available bandwidth, for example by invoking a scalability functionality of the underlying codec.

In summary, it is the intention of the content provider and of all users that the content provider can be trustworthily identified as the originator of each distributed substream of the media asset and that if multiple substreams are jointly consumed, users can verify their coherence, in order to for example guarantee that an audio stream was indeed intended to be associated to a video stream by the content provider.

Note: To address the above use cases, the standard requires carriage of embedded data in the media stream(s). Therefore, discovering if a media asset is fake or not, without carriage of the required embedded data is out of the scope of the solutions.

# System level requirements

## General aspects

The requirements listed fall within the scope of metadata embedding and referencing for verifying the media authenticity.

Solutions need to fulfill generic requirements for media streaming, such as dealing with random access into the streams, dynamic adaptive switching and content splicing.

In order to meet the requirements listed below for a single compressed version of the multimedia asset, it is desirable that the underlying compression standards like MPEG-D USAC or VVC both support a trustworthy verification mechanism that can be carried out independently (supporting important functionalities of the standards) but at the same time link such a verification mechanism to a joint verification method for substreams associated to different media types/codecs. Such a joint verification mechanism may be supported by the relevant media data at the system level. Therefore, requirements are developed that require solutions to be transported in media bitstreams and system streams to enable media authentication.

## System level specific requirements

Based on the identified use cases, the following requirements have been identified.

1. The solution shall support the verification of media content encapulated in ISOBMFF or delivered with MPEG DASH.
2. The solutions shall provide means to **embed authenticity information** into media assets.
   1. A media asset may consist of several media types and each media type may have one or more media tracks and potentially along with the associated DASH manifest. One or more media types also may be encapsulated in one media track. The solutions shall be flexible enough to embed such information at the elementary streams, as well as the system layer(s) to signal the cross-authentication of one or more elementary streams.
   2. Trust verification of several media types in separate tracks/streams shall ensure that the correlation and synchronization between different streams can be verified.
3. The solution shall **support digital signing** of media assets, metadata and content by producer(s) to establish authenticity and integrity of the content by:
   1. Supporting various **hashing** methods.
   2. Supporting **verification** of each media component by information embedded in the media component itself.
   3. Supporting joint **verification** of more than one media component associated with each other.
4. The solution shall support the carriage of the **provenance information** and/or the binding data to the provenance informationwith the media content. The solution shall be flexible enough to support various provenance schemes (e.g. schemas required by different regulations).
5. The solutions shall provide means to allow for **random access** into the media tracks/streams while, at the same time, allowing for **verification of temporal coherence** between random access segments.
6. The solution shall provide means to allow the **dynamic switching** of the media tracks/streams, when applicable, e.g. in the case of streaming, switching between different representations of the same media component shall allow the verification of the media component as a single part of a media asset.
7. The solution shall provide the possibility for **the verification of different time intervals** of the media assets, i.e. the authentication of each time interval of a media asset may need to be done independently of the other periods of the media.
8. The solution shall allow the verification of media assets that are captured and delivered in real time (i.e. **live**) as well as for later delivery (i.e. **on-demand**). It should offer configurable latency for the authentication process. The viewer shall be able to verify the content as being delivered in real time regardless of whether it was produced in real time or post-processed after production.
9. The solution shall allow the verification of the content distributed via broadcast or other means of one-way distribution where authentication information (e.g. keys) are exchanged via side channels (e.g. using secure on-device storage or other   
   “offline” methods).
10. The solution shall provide means to **modify specific parts of the media streams** while still allowing verification of authenticity for media assets by:
    1. Supporting **scalability** when codecs allow for it.
    2. Support **multiple component** coding (e.g. multi-view or depth-map, texture, geometry, etc.)
    3. Supporting **authoring of secondary information and metadata** (such as MPEG-H 3D audio user interaction and DRC metadata) in a downstream device.
    4. Support for media assets that are stored in single **or multiple files**, where applicable.
    5. Support of **late binding** between different media components or media objects.
11. The solution should, as much as possible, provide some **means of robustness** against loss such as video frame loss and/or audio dropouts, allowing partial verification of more important parts in cases of partial loss.
12. The solution should provide means to verify different substreams of a media track/stream with different verification and/or hashing methods, where applicable.
13. The solution shall provide means to embed authenticity information into media tracks and/or associated DASH manifestsin a way that does not have a significant impact on the bitrate (average, peak) of the media bitstream as much as possible.

# Gap Analysis

The following is a summary of contributions discussed at MPEG#151, Daejeon, meeting.

### [m73494 Analysis of systems media authenticity requirements based on MPEG-7 and MPEG-21](https://git.mpeg.expert/MPEG/Systems/explorations/-/issues/24)

MPEG 7 and MPEG 21 need extensions to be useful for media authentication and since they are not widely adopted, it is recommended to start an activity to extend ISOBMFF and DASH independent from them. The assumed scope is not to create a new provenance format but support of carrying provenance information in those standards.

### [m73485 Gap analysis of ISOBMFF and DASH to support media authentication requirements](https://git.mpeg.expert/MPEG/Systems/explorations/-/issues/23)

Several gaps are identified for ISOBMFF and DASH by going through the requirements and discussing what needs to be used in ISOBMFF and/or DASH and what is missing.

### [m73486 Gap analysis of C2PA to support media authentication requirements in ISOBMFF and DASH](https://git.mpeg.expert/MPEG/Systems/explorations/-/issues/25)

A gap analysis of using of C2PA with ISOBMFF and DASH is provided. There were different views at the meeting on how C2PA can be used with ISOBMFF and DASH.

## [m73486 Gap analysis of C2PA to support media authentication requirements in ISOBMFF and DASH](https://git.mpeg.expert/MPEG/Systems/explorations/-/issues/25)

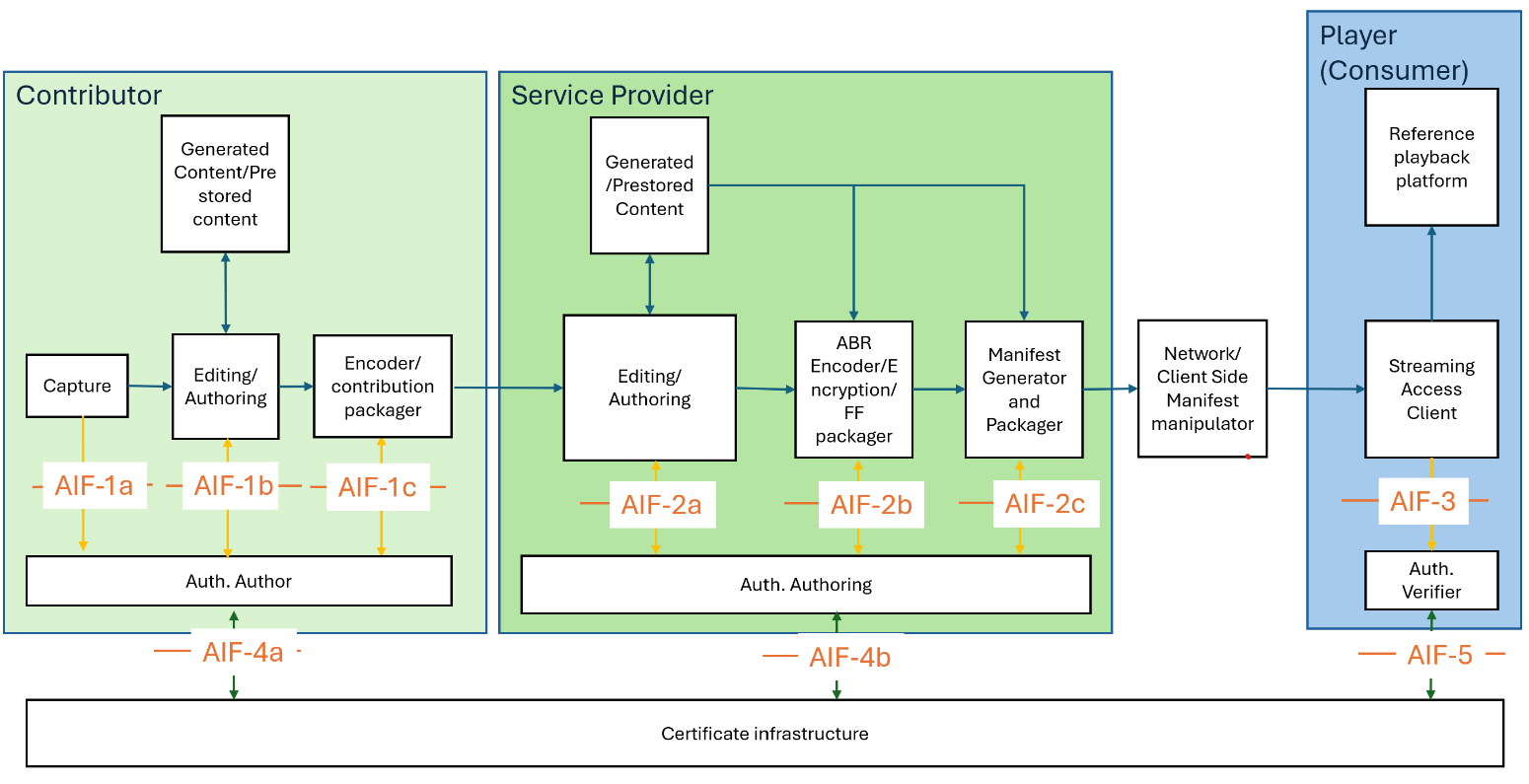
C2PA for live streaming has challanges for temporal binding of segments. A solution is proposed based on DSC VSEI and to carry the information at the system layer.

# Workflows (based on m74553 and m7440)

## Streaming Workflow (m74553)

<https://git.mpeg.expert/MPEG/Systems/explorations/-/issues/32>

Figure 1 defines a general end-to-end workflow for media authentication of media streaming applications.



**Figure 1: The media authentication workflow for streaming**

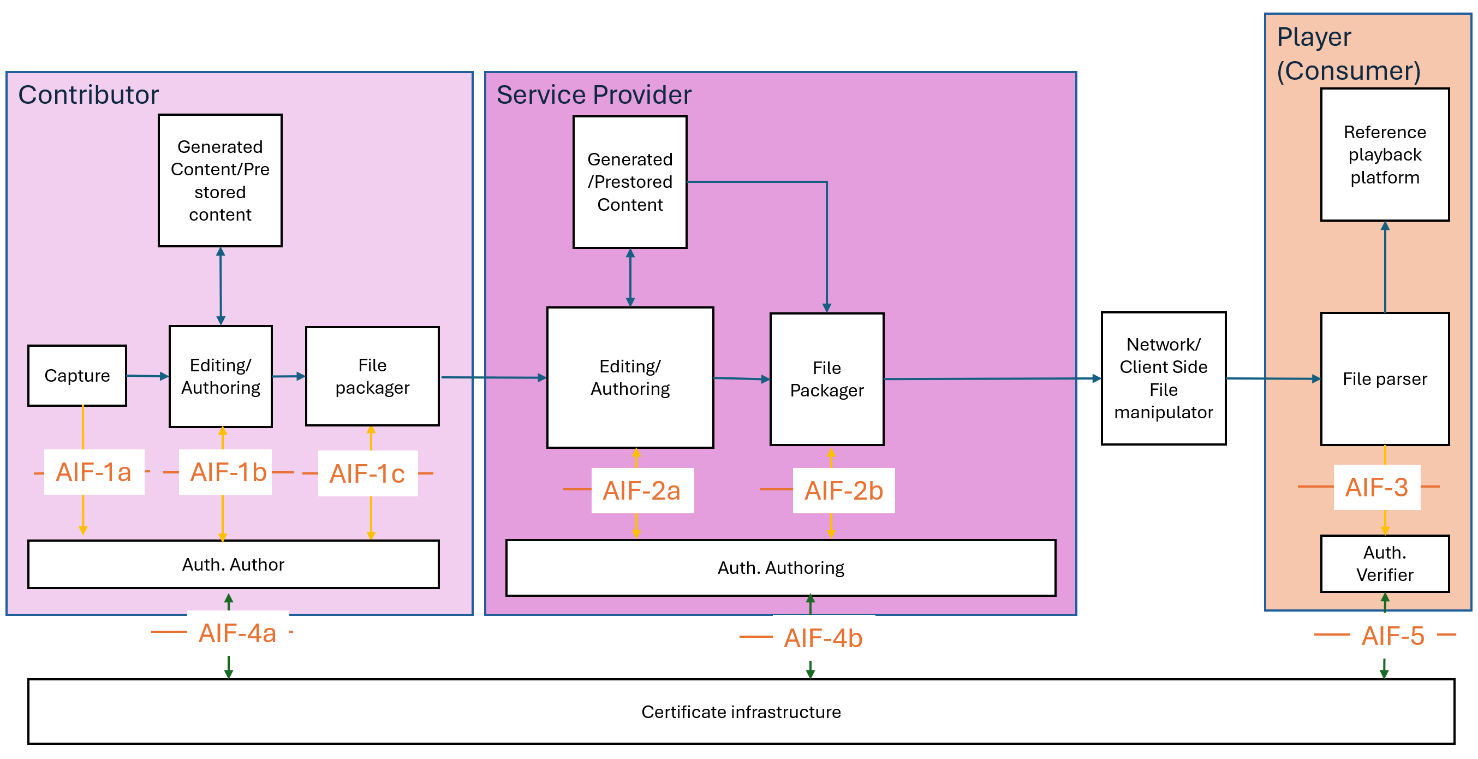
In this figure, the workflow consists of:

1. Media is captured by the device. Raw authentication information is provided to the authentication creator (AIF-1a).
2. The content is edited, manipulated and mixed with prestored and/or generated content. The provenance information is updated (AIF-1b)
3. Media is encoded and packaged. The authentication creator runs algorithms on the data (AIF-1c), interacts with the certificate infrastructure if needed (AIF-4a) and includes the authentication metadata in encoded and packaged data (AIF-1c). Optionally, it may include storing some data in the cloud (AIF-4a).
4. Media is provided to the content /service provider.
   1. The content received may be verified by the content/service provider (AIF-2a).
   2. The authentication metadata may be removed/redacted by the content/service provider.
5. The content is edited and revised. It may be manipulated, enhanced or revised and/or combined with other pre-stored or generated content. The provenance information is provided to the authentication manipulation/authoring module along with the provenance information provided at the contribution (AIF-2a). The new parts of the content may have their own provenance information.
6. Additional authentication metadata is included during the media packaging and manifest generation (AIF-2b/2c).
7. The media presentation may be repackaged at the network or client side or transmitted by a transmitter. Such operations maintain the authentication metadata for the part of the media originated from the service provider.
   1. The inserted content (such as inserted ads) should not impact the authentication metadata nor the verification of the main content.
   2. Each inserted content may add its own authentication information.
   3. If the inserted content causes failure in the verification of the main content, it should either
      1. Remove signaling that the manifest to avoid confusing the player, or
      2. Add the new authentication metadata for the combined content.
8. The player/receiver uses the verification module (AIF-3) for media verification through AIF-5.
   1. The player may use the manifest to select the media that can be verified.
   2. The verification may occur during the playback of media.

## File-based workflow (m74553)

<https://git.mpeg.expert/MPEG/Systems/explorations/-/issues/32>

Figure 2 describes the workflow for file-based transfer applications.



**Figure 2: The media authentication workflow for file-based transfer**

In this figure, the workflow consists of:

1. Media is captured by the device. Raw authentication information is provided to the authentication creator (AIF-1a).
2. The content is edited, manipulated and mixed with prestored and/or generated content. The provenance information is updated (AIF-1b)
3. Media is encoded and packaged into one or multiple files. The authentication creator runs algorithms on the data (AIF-1c), interacts with the certificate infrastructure if needed (AIF-4a) and includes the authentication metadata in encoded and packaged data (AIF-1c). Optionally, it may include storing some data in the cloud (AIF-4a).
4. Media is provided to the content /service provider.
   1. The content received may be verified by the content/service provider (AIF-2a).
   2. The authentication metadata may be removed/redacted by the content/service provider.
5. The content is edited and revised. It may be manipulated, enhanced or revised and/or combined with other pre-stored or generated content. The provenance information is provided to the authentication manipulation/authoring module along with the provenance information provided at the contribution (AIF-2a). The new parts of the content may have their own provenance information.
6. The media presentation may be repackaged at the network or client side or transmitted by a transmitter. Such operations maintain the authentication metadata for the part of the media originated from the service provider.
   1. The inserted content (such as inserted ads) should not impact the authentication metadata nor the verification of the main content.
   2. Each inserted content may add its own authentication information.
   3. If the inserted content causes failure in the verification of the main content, it should either
      1. Remove signalling that the manifest to avoid confusing the player, or
      2. Add the new authentication metadata for the combined content.
7. The player uses the verification module (AIF-3) for media verification through AIF-5.
   1. The verification may occur during the playback of media.

The media authentication interfaces are a subset of the ones identified for the streaming workflow. All interfaces are the same except AIF-2c, which does not exist in this case.

## Cross authentication in the file-based delivery workflow (m74440)

<https://git.mpeg.expert/MPEG/Systems/explorations/-/issues/34>

Figure 3 is the file-based workflow (originally presented in the contribution m73957 [7]) modified here to highlight where cross-authentication (X-auth) mechanisms may be applied.

A diagram of a computer

AI-generated content may be incorrect.

Figure 3 File-based workflow with cross authentication (X-Auth) of different Media types

On the contributor side, media components possibly with different media types (e.g. video, audio tracks, subtitles, etc.) of a media asset are captured (original content), retrieved or generated. Each component is authenticated individually through the AIF-1a and AIF1b interfaces. Once authenticated, all media components with different media types building the media asset are packaged with cross-authentication performed through the AIF-1c interface. The resulting packaged file is then made available to the service provider. There can multiple contributors providing media components with different variants or different media types to the Service provider e.g. different audio tracks, etc.

AIF-5

On the service provider side, the packaged file can be edited by modifying existing media types, reusing prestored content or incorporating new ones (possibly from multiple contributors). Similar to contributors, the service provider must ensure cross-authentication of the new packaged file, handled through the AIF-2a and AIF-2b interfaces.

## Content properties (m74553)

<https://git.mpeg.expert/MPEG/Systems/explorations/-/issues/32>

The following content properties are considered:

1. The content may be produced for on-demand, live or low-latency live consumption.
2. The content may be produced as single or multi-period content by the contributor and further be spliced to multi-period content by the provider.
3. Time-based events may be added by the contributor or service provider to the content.
4. The content may be manipulated by the provider or by the client to be spliced with other content (such as ads). Use cases such as DASH MPD insert event, with dynamic ad duration and early termination, are considered.

## Authentication properties (m74553)

<https://git.mpeg.expert/MPEG/Systems/explorations/-/issues/32>

The following authentication properties are considered:

1. The authentication metadata provides information to verify the content for specific verification periods, i.e. each time interval of the content has authentication metadata.
   1. Since the content may be produced on-demand, live or in the low-latency case, the media verification periods are defined to enable each use case accordingly.
   2. If authentication verification requires insertion of period transition (i.e. changing the period structure), the resulting content shall conform to the DASH standard, including not losing the ad insertion opportunities.
   3. Adding ad insertion opportunities shall still allow verification of the content for each verification period intended by the service provider.
   4. A pass-through verifiable content is only possible if the content is contained in one or more periods, marked as “unchangeable”.
   5. Cross-media authentication, i.e. authentication between different tracks or representation of media components, needs to be addressed.
2. The manifest and its updates may provide:
   1. High-level information about the existence of authentication metadata and their format, and possible playback constraints.
   2. The cross-media authentication information, i.e. which tracks are expected to be authenticated together.

## Interfaces (m74553)

The following common interfaces are identified in the above workflows:

* AIF-1a: capture provenance data
* AIF-1b: authentication data created during editing
* AIF-1c: authentication data embedding in encoding/packaging
* AIF-2a: provenance data update during authoring
* AIF-2b: authentication data embedding in the media content
* AIF-2c: authentication data embedding in manifest/packaging
* AIF-3: extraction of authentication data for verification
* AIF-4a/4b: auth. author cloud interface
* AIF-5: auth. verifier cloud interface

## Authentication functions (m74553)

### Authentication author

* Validate the previously authenticated media
* Receive capturing data and create provenance information
* Create and update the provenance information
* Calculation of hashes and signatures.

### Authentication verifier

* Verification of the previously authenticated media

## Interfaces’ features (m74553)

### AIF-1a

#### Capture-> Auth. Author

* The capturing provides the authentication authoring module with the characteristics of the capture information, such as time, location, and format. This interface is out of the scope of MPEG ISOBMFF and DASH work.

### AIF-1b

#### Editing-> Auth. Author

* Provenance information of the generated/prestored content along with its timing information

### AIF-1c

#### Encoder/Packager-> Auth. Author

* Encoded media tracks

#### Auth. Author-> Encoder/Packager

* Provenance information packaged in ISOBMFF boxes
* Signaling authentication in various tracks/track fragments
* Signaling authentication information in the DASH manifest

### AIF-2a

#### Encoder/Packager-> Auth. Author

* Authentication information, including provenance information of
  + Received content
  + Generated/Prestored content

#### Auth. Author-> Encoder/Packager

* Verification results of the received content

### AIF-2b

#### Encoder/Packager-> Auth. Author

* Encoded media tracks

#### Auth. Author-> Encoder/Packager

* Provenance information packaged in ISOBMFF boxes
* Signatures
* Signalling of the authentication in various tracks/track fragments

### AIF-2c

#### Packager-> Auth. Author

* Packaging information

#### Auth. Author-> Encoder/Packager

* Signaling of the authentication information for the DASH manifest

### AIF-3c

#### DASH Access client-> Auth. Verifier:

* Authentication info at manifest for selection
* Media tracks/segments for verification

#### Auth. Verifier-> DASH Access client

* Selection of the adaptation sets that can be verified
* Verification results

# Potential work items (m74553)

<https://git.mpeg.expert/MPEG/Systems/explorations/-/issues/32>

## ISOBMFF

### System-level authentication information in tracks/track fragments

* There is no standard mechanism to include the authentication metadata for each track/track fragment and cross authentication info (between tracks) in ISOBMFF.
* Currently C2PA uses uudi/udta to include c2pa information (as a JUMBF box). In this case, the cross-authentication information exists in the C2PA manifest. Using the ISOBMFF dedicated first-class citizen box to signal the media authentication scheme and its parameters (including C2PA) is a more preferred solution since the discovery and playback of that box would be more universally supported. Such signaling should cover multi-rate encoding of the same media component as well as the cross media component authentication information.
* Adding a signaling mechanism at the file format level is useful to indicate the authentication information of various tracks and how they are related to each other, enabling a parser to identify the correlation between tracks before invoking the authentication verification and simplify AIF-3.

## Elementary authentication information in tracks

* There is no standard mechanism to signal the existence of authentication and its characteristics of elementary streams at the track level.
* There is no mechanism of signaling cross-track authentication for elementary authentication information.
* There is no mechanism to tie the elementary stream authentication to the system layer authentication.

## DASH

### Signaling at the manifest

* The DASH standard does not support a standard way to signal the existence of the authentication information in the manifest. The signaling can be used for track selection as well as to simplify the interaction between the DASH access player and authentication verifier. The signaling can include signaling of the scheme and exposing high-level information of the scheme, as well as the relationship between the media authentication of various adaptation sets.

# Proposed candidate solutions

## Media cross authentication signaling in ISOBMFF (m74440)

<https://git.mpeg.expert/MPEG/Systems/explorations/-/issues/34>

A first approach may focus on file-based media assets, containing multiple authentic media types as previously described.

This would involve defining a new file-level container that:

* describes the authentic combination of these media types, and
* includes all the information necessary to verify the authenticity of the grouping.

As an example, if we consider the ISO Base Media File Format (ISO/IEC 14496-12, see [8]), no such container is specified for authentication purposes. However, a grouping mechanism is already available through the use of EntityToGroupBox, which allows the description of a set of items and/or tracks along with their grouping parameters.

aligned(8) class EntityToGroupBox(grouping\_type, version, flags)

extends FullBox(grouping\_type, version, flags) {

unsigned int(32) group\_id;

unsigned int(32) num\_entities\_in\_group;

for (i = 0; i < num\_entities\_in\_group; i++) {

unsigned int(32) entity\_id;

}

// additional data may be specified depending on grouping\_type

}

This EntityToGroupBox could be extended to explicitly define associations among entities within the file to ensure that this grouping is authentic and reliably verified.

## Signaling in MPEG-2 TS (m74512)

<https://git.mpeg.expert/MPEG/Systems/explorations/-/issues/36>

M74512 proposes a dedicated descriptor to signal information regarding authenticity for a program and its elementary streams within the Program Specific Information (PSI) that is part of a Transport Stream according to ISO/IEC 13818-1 MPEG-2 systems. It suggests specifying a descriptor including fields that support the joint evaluation of authenticity of different elementary streams, i.e. the detection if an authenticated elementary stream was changed or if there is a mismatch between different elementary streams.

# Discussions at MPEG152

* There are existing tools that can be used, such as the signaling SEI prefixes.
* Specific mechanism may not be an issue, but finding the information in the system layer is important.
* Understanding the granularity and sequencing of data operation in this context would provide more insights for the design goals.

# Recommendations

The following steps are recommended:

1. Continue the gap analysis of the existing MPEG ISOBMFF and DASH standards to address the requirements in this document. .
2. Continue the gap analysis of other MPEG standards and non-MPEG standards for addressing the requirements in this document when used in conjunction with MPEG ISOBMFF and DASH.
3. Define media authenticity workflows for use cases identifying the process by various actors and entities, and required functionalities and interfaces between them.
4. Investigate if each of the following standards requires a generic solution for carriage and signaling of the media authentication and provenance metadat:
   1. ISOBMFF/CMAF
   2. DASH
   3. MPEG 2TS
5. Collaborate with the relevant organizations such as C2PA, EBU and SVTA/DASH-IF on the media authentication

# References

1. UNESCO: <https://en.unesco.org/fightfakenews>
2. ISO/IEC JTC 1/SC 29/WG1, “Use cases and requirements for JPEG Trust”, N100886, Sapporo, 13-19 July 2024.
3. C2PA Technical Specification, <https://c2pa.org/specifications/specifications/2.1/specs/C2PA_Specification.html>
4. JPEG Trust, https://jpeg.org/jpegtrust/documentation.html