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

Text

Description automatically generated ISO/IEC JTC 1/SC 29/WG 03 N1516

**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:** 2025-04-04

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

**No. of pages:** 10 (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** **1516**

**APR 2025, Online**

|  |  |
| --- | --- |
| **Source** | **MPEG Systems working group** |
| **Status** | **Output** |
| **Title** | **Exploration on media authenticity and provenance indication with the MPEG Systems technologies** |
| **Serial Number** | **25052** |

INTERNATIONAL STANDARD© ISO/IEC 2015 – All rights reservedISO/IEC 15444-12:2015(E) 63Part 12: ISO base media file formatInformation technology — JPEG 2000 image coding systemTechnologies de l'information — Codage des objets audiovisuels — Partie 12: Format ISO de base pour les fichiers médiasInformation technology — JPEG 2000 image coding system — Part 12: ISO base media file formatE2015-02-20(60) PublicationISO/IECISO/IEC J   International Standard 2015ISO/IEC 15444‑ISO/IEC 15444‑12ISO/IEC 15444-12  Coding of audio, picture, multimedia and hypermedia informationInformation technology 291 2見出し 2見出し 1    02 STD Version 2.1c260   4C:\Users\shinji\_w\AppData\Roaming\Microsoft\Templates\STD\w15177\_14496\_5th.-restyle-R1.doc

# 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.

# Scope of this exploration

The MPEG Systems standards are widely used for the delivery of the media. Therefore, the goal of this exploration is to investiage to what extend the MPEG System standards (in particular ISOBMFF and DASH) address the relevant media authentication and AI generated/altered content marking use-cases by the use of the existing tools, and/or by combining them with the non-MPEG standards. The system level requirements are identified in the following clause.

The following tasks are identified to achieve the goals of this exploration:

1. Study the existing MPEG ISOBMFF and DASH standards and identify the subset of requirements of 1 that are already addressed by these standards, and identify gaps.
2. Study the existing non-MPEG standards, and identify the subset of requirements of 1 that are already addressed by these standards, and identify gaps.
3. Solicit initial candidate solutions for addressing these gaps.

# 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