MULTIDISCIPLINARY RESEARCH INSTITUTE INDIA

Multidisciplinary Journal of Research in Engineering and Technology

ISSN: 2348-6953

Volume 13 Issue 02 (2026) | Pages: 314-320

Digital Provenance: Cryptographic Content Authentication

Minakshi M. More[1], Vishwanath Hatti[2], Anis Khan[3], Krishnakant Gupta[4], Anurag Kumar Goutam[5], Yash Gurharikar[6]

[1,2,3,4,5,6]Department of MCA, MES' IMCC, Pune, Maharashtra,

[1]*mst.imcc@mespune.in*, [2]*vhatti14@gmail.com*, [3]*aniskhan20171@gmail.com*, [4]*krishnakant.97200100@gmail.com*, [5]*anuraggoutam133@gmail.com*, [6]*yashgurharikar420@gmail.com*

Peer Review Information

***Type**: Article*
***Received**: 20 March 2026*
***Revised**: 03 April 2026*
***Accepted**: 21 May 2026*
***Published**: 03 June 2026*

Abstract

The modern digital landscape is currently navigating what Nadia Naffi (2025) describes as a "crisis of knowing"—a threshold where human senses are no longer sufficient to distinguish between a genuine record of history and an AI-generated lie. [1] With annual fraud losses projected by the World Economic Forum (2025) to reach 37,60,42,20,00,000.00 Indian Rupee by 2027, our society is facing an existential threat to its shared reality. [2] This paper moves beyond the flawed "Detect and Debunk" model and proposes a theoretical shift toward a "Prove and Protect" framework centered on Digital Provenance. We analyze the structural failures of legacy metadata systems identified by ISACA (2025) and the "Middle Mile" stripping of data by social platforms documented by Kaptur (2024). [3] In response, we propose the "Smart Camera Button" architecture, inspired by the "Signing Right Away" (SRA) model developed by Yejun Jang (2025). [4] This framework enforces mandatory silicon-level signing at the "First Mile" of content creation. By isolating the imaging pipeline within a Secure Enclave (Apple, 2025) and leveraging C2PA (2025) standards, we can rebuild trust through verifiable, mathematical history. [5]

Keywords: Digital Provenance; Cryptography; Deepfakes; Secure Enclave; Coalition for Content Provenance and Authenticity; Content Authentication.

How to Cite This Article

More, M. M., Hatti, V., Khan, A., Gupta, K., Goutam, A. K., & Gurharikar, Y. (2026). Digital provenance: Cryptographic content authentication. *Multidisciplinary Journal of Research in Engineering and Technology*, 13(2), 314–320.

Introduction

For over a century, a photograph was considered a mechanical witness to the truth. As Andreas Kaufmann of Leica (2023) notes, cameras have historically stood witness to iconic moments in world history [8]. But today, that mechanical trust has evaporated. We have entered an era of *"digital ghosts,"* where hyper-realistic images and videos are created from nothing but code. This is not just a problem for photographers; it is a serious challenge for journalism, law, and our shared perception of reality.

The Problem: The Gospel of the Pixels

The internet currently operates on what can be described as the *"gospel of the pixels"*—if the dots on your screen appear real, they are accepted as truth. However, as GAFA (2025) research demonstrates, this assumption is no longer valid. With modern AI systems capable of cloning a voice from just a few seconds of audio or performing real-time face swaps in video calls, visual and auditory realism can no longer be trusted [9]. As a result, we are increasingly vulnerable, relying on identification methods that are fundamentally flawed.

The Goal: Notary, Not Detective

As outlined in the C2PA Technical Specifications (2025), the solution is not to improve detection of fake content, but to fundamentally change how authenticity is established [6]. Instead of acting as *"detectives"* attempting to identify every forgery, systems should function as *"notaries"* that verify genuine content. The objective is to establish Digital Provenance—a verifiable digital "birth certificate" that provides an unbroken chain of custody from the moment an image is captured to the point it is viewed.

Existing Theories and Literature Review

The Hidden Mess of Metadata

Most people don't realize that every time they capture a photo, their device is simultaneously storing additional information within the file known as *metadata* (such as EXIF data). This metadata acts like a hidden diary entry, recording technical details such as the date, GPS coordinates, and camera model. While this appears to be a reliable way to verify authenticity, research by ISACA (2025) highlights that metadata suffers from two critical weaknesses [4].

The "Post-it Note" Vulnerability

Michael Steidl of the IPTC (2024) describes standard metadata as being as fragile as a post-it note attached to a box [3]. Anyone with basic tools can remove or modify this information with ease. For example, a malicious user can alter a photo's GPS coordinates to falsely indicate that it was taken in a conflict zone, when in reality it was captured elsewhere. Standard software lacks the ability to detect such tampering, making metadata unreliable as proof of authenticity.

The "Treasure Map" Theory

From a technical perspective, metadata often relies on "pointers," which act as a map directing software to where information is stored within a file [4]. According to Kaptur (2024), when a photo is edited and saved using basic software, these pointers are often removed to reduce file size [3]. As a result, the metadata may still exist within the file, but it becomes inaccessible. This is comparable to having a treasure map where the paths have been erased—making it impossible to locate the information.

The "Social Media Shredder"

Even when users do not intentionally alter metadata, digital platforms contribute to its loss. Research by Imatag (2024) indicates that up to 80% of images online have lost their original metadata [3]. When images are uploaded to platforms such as Facebook or WhatsApp, they are compressed and re-encoded. During this process, most embedded data is stripped away, leaving the image without any verifiable history. This creates an environment where manipulated and authentic content become indistinguishable.

The "Liar's Dividend"

This erosion of trust has led to what Bobby Chesney and Danielle Citron (2024) describe as the *"Liar's Dividend"* [1]. As awareness of deepfakes increases, individuals can dismiss genuine evidence—such as videos of misconduct—by claiming it is fabricated. When the authenticity of all content becomes questionable, even truth loses its value. This has already resulted in significant financial and social consequences, including the ₹2,45,36,75,355 Hong Kong transfer fraud reported by **GAFA (2025)** [9].

System Architecture Overview

The proposed Digital Provenance system can be understood as a layered architecture consisting of three primary stages: capture, sealing, and verification.

1. Capture Layer: At the moment of image capture, raw sensor data is transmitted through a secure hardware channel. Unlike traditional pipelines, this prevents any software-level interference before authentication.
2. Cryptographic Sealing Layer: Within the Secure Enclave, the image undergoes hashing and digital signing. This process generates a unique cryptographic identity tied to both the content and the device.

3. Storage and Distribution Layer: The signed image is stored along with its provenance metadata. Even when shared across platforms, the cryptographic signature remains embedded within the content.
4. Verification Layer: At the point of consumption, applications such as browsers or verification tools use public keys to validate authenticity. This ensures that trust is established at the time of viewing rather than creation.

This layered architecture enforces security at multiple stages, significantly reducing the attack surface and improving overall system reliability.

Proposed Framework: The "Smart Camera Button"

We propose establishing a *Hardware Root of Trust* that begins at the First Mile—the exact microsecond when light hits the camera sensor [5].

The Secure Pipeline Logic (The Inventive Step)

In current systems, a camera captures an image, stores it, and only later attempts to apply a digital signature through software. This creates a vulnerability window where an attacker could inject a manipulated file.

The proposed "Smart Camera Button" architecture, inspired by the "Signing Right Away (SRA)" model developed by Yejun Jang (2025), eliminates this gap by enforcing a hardware-controlled, mandatory pipeline [5].

1. Direct Silicon Tunneling
Raw pixel data is transmitted through a private, encrypted hardware channel (using the MIPI CSI-2 protocol) directly from the camera sensor to the Secure Enclave [5].

2. The Silicon Seal
Within the Secure Enclave, described by Apple (2025) as an isolated security processor, a unique digital fingerprint (hash) of the image is computed instantly [7].

3. The Fused Identity
The system signs this hash using a Private Key that is permanently fused into the device hardware during manufacturing. This key cannot be extracted, copied, or accessed by applications.

4. Mandatory Storage Sealing
The storage system is designed to reject any image that does not contain a valid cryptographic signature. This ensures that only authenticated images are stored in the device gallery [5].

How the Technology Works (Simple Analogies)

- Hashing (The Coffee Grinder): Hashing works like a grinder. You input a specific set of coffee beans (the photo), and it produces a unique powder (the hash). As explained by Professor Messer (2025), this process is one-way—once ground, the beans cannot be reconstructed. Even a single pixel change results in a completely different output.
- Digital Signatures (The Wax Seal): The camera applies a "digital wax seal" to the hash using its private key. The viewer (e.g., a browser) uses the manufacturer's public key to verify whether this seal has been tampered with.

Technical Supplement: The Logic of the Digital Seal

To maintain clarity, we examine the mathematical foundation of **SHA-256 hashing** and **ECDSA signatures** used in this framework.

The Hash Function (H)

A photo *(P)* is converted into a hash *(h)*:

$h = H(P)$

Even a minor modification *(ΔP)* produces a completely different result:

$H(P + \Delta P) \neq H(P)$

The Signature (S)

The Secure Enclave uses the private key *(d)* and the hash *(h)* to generate a signature:

$S = sign(d, h)$

The Verification (V)

A browser uses the public key *(Q)* to retrieve the original hash:

$h_{orig} = verify(Q, S)$

If h_{orig} matches the hash of the current image, the content is authentic. Since the private key is permanently embedded in hardware, forging a valid signature is computationally infeasible—even with advanced computing resources.

Fig. 1. *Workflow of Digital Seal*

Discussion: Why It Is Better

This hardware-first approach introduces multiple *layers of truth* that cannot be replicated or forged through software-based manipulation.

1. Proving the World is 3D
A common method used in deepfake attacks is capturing a photo of a screen displaying manipulated content. This technique attempts to bypass detection by presenting fake content as if it were real.

However, the Sony Alpha Universe (2025) model incorporates 3D depth information directly into the cryptographic seal [12]. The camera sensor verifies that incoming light originates from a real three-dimensional environment rather than a flat two-dimensional display.

Because this depth data is embedded within the signed record, it becomes mathematically verifiable. As a result, a viewer can confidently determine that the image was captured from a real, physical subject rather than a replayed or synthetic source.

2. Surviving the "Social Media Shredder"
A major challenge in maintaining digital authenticity is the way platforms process uploaded media. Social media services often compress and re-encode images, stripping away embedded metadata in the process.

To address this, Collomosse et al. (2024) propose the concept of Durable Content Credentials [11]. This approach embeds an invisible watermark directly into the image pixels.

Even if metadata is removed, a verification system can scan the image, extract the watermark, and retrieve the original signed provenance data from a secure cloud registry. This ensures that authenticity can be preserved even after aggressive platform-level transformations.

Threat Model and Security Analysis

To evaluate the robustness of the proposed Digital Provenance framework, it is essential to analyze potential attack vectors and how the system mitigates them.

Adversarial Model

We consider an attacker with the following capabilities:

- Ability to manipulate image pixels using advanced AI tools
- Access to standard editing software capable of modifying metadata
- Ability to intercept and redistribute media content across platforms

However, the attacker does **not** have access to:

- The hardware-isolated Secure Enclave

- The device-specific Private Key

Attack Scenarios and Mitigation

1. Pixel Manipulation Attack

Any modification to the image changes its hash value:

$H(P') \neq H(P)$

This immediately invalidates the digital signature, making tampering detectable.

2. Metadata Forgery Attack

Unlike traditional EXIF metadata, provenance data is cryptographically bound to the image. Therefore, altering metadata does not affect verification unless the signature is recomputed—which is infeasible without access to the private key.

3. Replay Attack

An attacker may attempt to reuse a valid signature on a different image. However, since the signature is tightly coupled with the original image's hash, verification will fail.

4. Man-in-the-Middle Attack

Even if content is intercepted during transmission, any modification will break the signature chain, ensuring that integrity violations are detected.

Security Guarantees

The proposed system provides the following guarantees:

- Integrity: Any modification to the content is detectable
- Authenticity: The origin of the content can be verified
- Non-repudiation: The creator cannot deny ownership of the content

Implications and Applications

By 2026, this technology is transitioning from high-end cameras such as the Leica M11-P into devices used by the general public [8]. According to Trufo (2025), the Google Pixel 10 is expected to be the first mainstream smartphone to automatically sign every captured image using its internal Titan M2 security chip [13].

This advancement has significant implications across multiple domains:

- Journalism: As suggested by Nikon (2025), journalists can provide a verifiable and tamper-proof chain of custody for visual evidence—from the moment of capture in a conflict zone to its publication [14]. This enhances credibility and trust in news reporting.
- Legal Systems: Digital evidence becomes significantly more reliable in legal contexts. Since the content is protected by a hardware-backed cryptographic signature, any tampering can be detected, ensuring the integrity of evidence presented in court [1].

Social and Regulatory Impact

The EU AI Act (2026) mandates that AI-generated content must be clearly labeled. The C2PA "CR" (Content Credentials) icon serves as a standardized technical solution for fulfilling this requirement [15]. This enables users to easily distinguish between authentic and synthetic media.

Limitations and Future Work

While the proposed framework provides strong guarantees of authenticity, several challenges remain.

Limitations

1. Hardware Dependency: The system relies on secure hardware components such as the Secure Enclave. Devices that do not include such hardware cannot participate in this trust model, limiting universal applicability.
2. Adoption Barrier: For Digital Provenance to be effective, it requires widespread adoption across device manufacturers, software platforms, and regulatory bodies. Without ecosystem-wide support, its impact remains limited.

3. Privacy Concerns: Embedding metadata such as device identity, timestamps, and location information may raise privacy concerns if not managed carefully. Proper safeguards are necessary to balance transparency with user privacy.

Future Work

Future research directions include:

- Zero-Knowledge Proofs: Enabling verification of authenticity without exposing sensitive metadata
- Decentralized Registries: Leveraging blockchain or distributed systems for immutable provenance storage
- Cross-platform Standardization: Ensuring interoperability across devices, platforms, and ecosystems
- Real-time Verification Tools: Developing browser or application-level tools for instant authenticity validation

Case Study: Deepfake Financial Fraud

A notable real-world example that highlights the urgency of Digital Provenance is the increasing use of AI-driven financial fraud. In a widely reported case, attackers used deepfake audio to impersonate a company executive, successfully convincing employees to transfer millions in funds.

In such situations, traditional verification mechanisms fail because the content appears authentic to human perception. The voice, tone, and context are convincing enough to bypass manual checks.

However, under the proposed Digital Provenance framework, any generated or manipulated media would lack a valid cryptographic signature issued by a trusted hardware source. This absence of a verifiable signature would immediately indicate that the content is untrusted.

As a result, organizations could flag such content as suspicious without relying on subjective human judgment. Instead, decisions could be based on objective, mathematical verification.

This case demonstrates that Digital Provenance is not merely a theoretical concept, but a practical necessity—particularly in high-risk domains such as finance, law enforcement, and national security.

Conclusion

We are not merely improving camera technology—we are redefining the foundation of trust in the digital age. The emergence of hyper-realistic synthetic media has exposed a critical weakness in current systems: the inability to reliably distinguish between reality and fabrication. As discussed throughout this work, approaches based on detection and debunking are inherently reactive and insufficient in a world where AI-generated content continues to advance rapidly.

The proposed “Smart Camera Button” framework represents a fundamental shift from skepticism to certainty—from questioning every image to mathematically proving its origin. By embedding a hardware root of trust directly at the point of capture and enforcing cryptographic sealing through secure enclaves and standardized provenance frameworks such as C2PA, this approach establishes an immutable chain of custody that cannot be altered without detection [6].

This transformation carries significant implications across multiple domains. In journalism, it restores credibility to visual evidence. In legal systems, it introduces a new category of tamper-evident digital proof [1]. At a societal level, it addresses the growing issue of the “liar’s dividend,” where genuine content is dismissed as fake [1]. Most importantly, it shifts the burden of proof—from individuals attempting to detect deception to systems that inherently guarantee authenticity.

Looking ahead, the widespread adoption of hardware-backed digital provenance will not eliminate misinformation entirely. However, it will create a clear and verifiable distinction between authenticated reality and unverified content. In such a landscape, trust will no longer depend on perception but on provable mathematical guarantees.

Ultimately, the goal is not to classify content as “good” or “bad,” but to ensure that its origin and history are transparent, verifiable, and secure from the moment of creation. As digital media continues to shape human understanding, rebuilding this trust infrastructure is not only a technical necessity but a societal imperative.

Beyond technical implementation, the success of Digital Provenance will depend on global collaboration among hardware manufacturers, software platforms, and regulatory bodies. Standardization efforts such as C2PA must evolve into universally accepted protocols, similar to HTTPS in web security. Only through such coordinated adoption can authenticity become a default property of digital content rather than an exception.

In the long term, Digital Provenance has the potential to redefine how humans interact with digital media. It introduces a paradigm in which trust is no longer assumed but mathematically verified. This shift will not only protect individuals and organizations from deception but also preserve the integrity of information in an increasingly synthetic world.

References

1. Naffi, N. (2025). *UNESCO: Deepfakes and the Crisis of Knowing*. [Online]. Available: *https://www.unesco.org/en/articles/deepfakes-and-crisis-knowing*
2. C2PA (2025). *Technical Specifications for Content Provenance v2.3*. [Online]. Available: *https://spec.c2pa.org/specifications/specifications/1.0/specs/C2PA_Specification.html*
3. Collomosse, J., et al. (2024). *Durable Content Credentials: Three Pillars of Provenance*. IEEE.
4. Jang, Y. (2025). *Signing Right Away (SRA): End-to-End Secure Imaging Pipelines*.
5. European Union (2026). *The EU AI Act: Article 50 Transparency Obligations*. [Online]. Available: *https://eur-lex.europa.eu/eli/reg/2024/1689/oj*
6. Sony Corporation (2025). *Camera Authenticity Solution*. [Online]. Available: *https://www.sony.net/cas/*
7. Macfilos (2023). *Leica M11-P: Image Authentication at Point of Capture*. [Online]. Available: *https://www.macfilos.com/2023/10/26/leica-m11-p-content-credentials/*
8. GAFA (2025). *Deepfake Fraud Case Studies: Corporate Risk Analysis*.
9. Kaptur (2024). *Structural Failure of Image Metadata: Why 80% of Data is Lost*. [Online]. Available: *https://kaptur.co/whos-behind-the-image/*
10. Trufo (2025). *Google Pixel 10 and the Mainstreaming of Digital Provenance*.
11. Apple Inc. (2025). *Platform Security Guide: The Architecture of the Secure Enclave*. [Online]. Available: *https://support.apple.com/guide/security/secure-enclave-sec59b0b31ff/web*
12. World Economic Forum (2025). *The Economic Costs of Synthetic Media*. [Online]. Available: *https://www.weforum.org/reports/global-risks-report-2025*
13. ISACA (2025). *EXIF Metadata: A Subtle Cybersecurity Risk for Organizations*. [Online]. Available: *https://www.isaca.org/resources/news-and-trends/industry-news/2025/what-to-know-about-exif-data-a-more-subtle-cybersecurity-risk*
14. Nikon USA (2025). *Nikon Authenticity Service: C2PA Implementation*. [Online]. Available: *https://www.nikonusa.com/en/learn-and-explore/a/products-and-innovation/content-credentials.page*
15. Ryan, S. (2026). *The Birthmark Standard: Hardware Roots of Trust*.
16. Minakshi S. Tumsare (2018). *Statistical Analysis to Compare RSA and AES128 Algorithms for PIN-type Message Transactions*. Journal of Advance Research in Dynamical & Control Systems, Vol. 10, Special Issue.