Signing Right Away

Yejun Jang

October 5, 2025

Abstract

The proliferation of high-fidelity synthetic media, coupled with exploitable hardware vulnerabilities in conventional imaging pipelines, has precipitated a crisis of trust in digital content. Existing countermeasures, from post-hoc classifiers to software-based signing, fail to address the fundamental challenge of establishing an unbreakable link to reality at the moment of capture. This whitepaper introduces Signing Right Away (SRA), a comprehensive security architecture that guarantees the provenance of digital media from "silicon to silicon to signed file." SRA leverages a four-pillar security model—Confidentiality, Integrity, Authentication, and Replay Protection—inspired by the MIPI Camera Security Framework, but extends its scope beyond the internal data bus to the creation of a cryptographically sealed, C2PA-compliant final asset. By securing the entire imaging pipeline within a Trusted Execution Environment (TEE), SRA ensures that every captured image and video carries an immutable, verifiable proof of origin. This provides a foundational solution for industries reliant on trustworthy visual information, including journalism, legal evidence, and insurance. We present the SRA architecture, a detailed implementation roadmap informed by empirical prototyping, and a comparative analysis that positions SRA as the essential "last mile" in the chain of content trust.

1 Motivation

The digital information ecosystem is at a critical inflection point. The confluence of two powerful trends—the democratization of hyper-realistic synthetic media generation and the persistent neglect of fundamental hardware security in imaging devices—has created an environment where the authenticity of visual content can no longer be taken for granted. This section establishes the urgent need for a new paradigm in content authentication by examining the dual nature of this threat and the inadequacy of current solutions.

1.1 The Erosion of Trust

The advent of sophisticated generative models has fundamentally altered the landscape of digital content creation. Technologies such as OpenAI's DALL-E series for images [1, 19, 20], Sora for video [3], and Google's AudioLM for audio [2] have made the generation of high-fidelity synthetic content accessible to a global user base with minimal technical expertise. What once required specialized skills and significant computational resources can now be accomplished with simple text prompts.

While these tools offer unprecedented creative potential, they also represent a potent new vector for disinformation and fraud. The ability to generate convincing but entirely fabricated visual "evidence" poses a direct threat to social cohesion, financial markets, and democratic processes. A stark illustration of this threat occurred in May 2023, when an AI-generated image depicting an explosion near the Pentagon was circulated on social media by a verified account [18]. The viral spread of this fabricated image triggered a brief but significant dip in the S&P 500 index, demonstrating the tangible, real-world consequences of synthetic media weaponized for informational warfare. This incident is not an isolated anomaly but a precursor of a new reality where the foundational assumption of "seeing is believing" is irrevocably broken. The core problem is the erosion of a shared visual reality, undermining the very basis of trust in digital communication.

1.2 The Hardware Attack Surface

Compounding the threat from generative AI is a critical, yet widely overlooked, vulnerability at the hardware level of nearly all modern imaging devices, from smartphones to professional cameras. The connection between the image sensor and the host system's Image Signal Processor (ISP) is typically managed by the MIPI Camera Serial

Interface 2 (CSI-2) protocol [17]. In its standard implementation, the raw bitstream of pixel data transmitted over this interface is unencrypted and unauthenticated.

This architectural flaw creates a gaping security hole. It allows for a straightforward man-in-the-middle attack, where a malicious hardware component can intercept the data path and inject a completely fabricated video feed, which the host system will accept as genuine data from its own camera sensor. This is not a theoretical vulnerability; it is a practical one, enabled by commercially available hardware. Devices marketed as "HDMI to CSI-2 adapters," often based on bridge chips like the Toshiba TC358743XBG, are sold openly and can be used to perform precisely this type of data injection attack [12, 22]. An attacker can connect any HDMI source—such as a computer playing a synthetically generated video—to one of these adapters, which then outputs a valid CSI-2 signal that can be fed directly into a device's motherboard, bypassing the camera entirely.

The synergy between these two threats creates a perfect storm for sophisticated deception. An adversary can first use a generative model to create a hyper-realistic fake video of a significant event. Next, they can use an off-the-shelf HDMI-to-CSI-2 adapter to inject this video feed into a legitimate, trusted device (e.g., a specific model of smartphone). The device's operating system, having no reason to distrust the data arriving at its ISP, would process the video as if it were captured by its own sensor. If the device is running a content authentication application, it could then proceed to generate a cryptographically valid signature and a C2PA manifest for the fake content. The final result is a piece of synthetic media bearing a legitimate digital signature from a trusted hardware source. This combined attack vector circumvents nearly all conventional methods of verification, demonstrating that any solution that does not address security at the hardware source is fundamentally incomplete.

2 Related Works

In response to the "fake data" problem, several countermeasures have been proposed and deployed. However, a systematic analysis reveals that each suffers from limitations that render them insufficient to counter the combined hardware and software threat vector described above.

2.1 Neural Network Classifiers

One approach involves training deep learning models to discriminate between real and synthetic content. While these classifiers can achieve some success under controlled conditions, they are inherently brittle. Their performance degrades significantly when faced with common transformations like compression, which is ubiquitous on the internet [21]. Furthermore, they struggle to generalize to content produced by novel generator architectures they were not trained on [5]. The very nature of Generative Adversarial Networks (GANs), which co-evolve a generator and a discriminator (a classifier), means that generative models are actively trained to defeat such detection methods, locking classifiers in a perpetual and likely unwinnable arms race.

2.2 Digital Watermarking

Digital watermarking involves embedding a hidden, noise-like signal into media to track its origin or identify it as synthetic. Technologies like Google's SynthID are a prominent example of this approach, designed to mark AI-generated content in a way that is imperceptible to humans but detectable by a specific algorithm [13]. While useful for copyright enforcement and providing a "label" for synthetic media, watermarking is not a robust security mechanism. Watermarks can be fragile and may be damaged or removed by subsequent image processing, compression, or even simple cropping. More importantly, watermarking is a post-hoc labeling technique; it does nothing to prevent the initial injection of unsigned, malicious data at the hardware level.

2.3 Software-Level Digital Signing (Standard C2PA)

The Coalition for Content Provenance and Authenticity (C2PA) has established a vital open standard for creating interoperable, cryptographically signed metadata that records the provenance of a digital asset [7]. A C2PA manifest can detail an asset's creation, authorship, and editing history, providing a transparent record for consumers [6]. However, when implemented purely in software on an untrusted device, the C2PA standard has a critical limitation: it can only attest to the provenance of the digital data it is given. A software-based C2PA tool signs a manifest that says, "This application, running on this device, received these pixels and is now signing them." It has no way of verifying that those pixels originated from the device's actual camera sensor. This creates a "garbage in, gospel out" scenario. If a fake video stream is injected via a hardware attack, a software C2PA implementation will

Solution	Mechanism	Point of Application	Resilience to Tampering	Primary Use Case
Neural Classifier	Statistical pattern recognition to detect artifacts of generation.	Post-Processing / Distribution	Low. Fails on novel generators and compressed content.	Detecting known types of synthetic media.
Digital Watermarking	Embedding a hidden signal (e.g., SynthID) into the media file.	Post-Processing / At Generation	Medium. Can be degraded by transformations and is not a security feature.	Labeling and tracking content (e.g., copyright).
Software C2PA	Cryptographic signature over a manifest of metadata.	Post-Capture / Software Layer	High (for the manifest), but Zero for the content's origin.	Interoperable provenance logging in untrusted environments.
SRA	**Hardware-secured capture pipeline with authenticated encryption and TEE-based signing.**	**Point of Capture (Sensor-to-SoC)**	**Very High. Protects against hardware injection and software compromise.**	**Establishing verifiable, hardware-rooted proof of origin.**

Table 1: Comparative Analysis of Content Authenticity Solutions

dutifully sign it, lending a false aura of authenticity to the fabricated content. The signature is cryptographically valid, but the underlying asset is a lie.

As summarized in Table 2.3, existing solutions fail to provide a complete, resilient defense because they do not operate at the true point of origin: the hardware interface between the sensor and the processor. This is the critical gap that the Signing Right Away architecture is designed to fill.

3 Signing Right Away (SRA)

To address the fundamental vulnerabilities outlined in the previous section, a new architectural approach is required. Signing Right Away (SRA) is a comprehensive security framework designed to establish an unbreakable chain of trust that extends from the moment photons strike the image sensor to the final, portable, and independently verifiable digital asset. It is not merely a signing protocol but an end-to-end secure imaging pipeline that integrates hardware-level protection with standardized, cryptographically-verifiable provenance records.

3.1 Core Principles

The foundational philosophy of SRA is to create an unforgeable, hardware-enforced cryptographic link between a physical event and its digital representation "right away" at the moment of capture. The architecture is built on the premise that trust must be established at the earliest possible point in the imaging pipeline and maintained through every subsequent stage of processing. By securing the data path from "silicon to silicon"—from the image sensor's silicon to the host System-on-Chip's (SoC) silicon—SRA eliminates the hardware attack surface that plagues conventional systems. This hardware-rooted trust is then extended to the final file, creating a complete "photon to file" chain of custody.

3.2 A Four-Pillar Security Model

The SRA architecture is structured around four essential security pillars, inspired by the robust design principles of the MIPI Camera Security Framework but applied to the broader goal of creating a verifiable asset. Each pillar addresses a specific threat to the integrity of the imaging pipeline.

- **Authentication:** The process begins with a mutual cryptographic handshake between the image sensor module and the host SoC. This step ensures that only genuine, authorized hardware components are part of the pipeline. It cryptographically verifies the identity of the camera, preventing the connection of rogue sensors or malicious data injection hardware, such as the HDMI-to-CSI-2 adapters previously discussed. This establishes the trusted endpoints of the secure channel.

- **Confidentiality:** Once trust is established, all image data transmitted over the MIPI CSI-2 bus is encrypted using a strong, authenticated stream cipher (e.g., AES-GCM). This renders the raw camera feed unintelligible

to any unauthorized party attempting to eavesdrop on the physical connection, protecting the privacy and content of the data while it is in transit.

- **Integrity:** Alongside encryption, every packet or frame of image data is accompanied by a Message Authentication Code (MAC). This cryptographic tag is calculated by the sensor and verified by the SoC. Any modification to the data in transit, no matter how small, will cause the MAC verification to fail. This guarantees that the data received by the SoC is bit-for-bit identical to the data that left the sensor, preventing any form of in-transit tampering or pixel manipulation.

- **Replay Protection:** To defend against an attacker capturing a valid, encrypted frame and re-injecting it into the stream at a later time, the protocol incorporates anti-replay mechanisms. This is typically achieved through the use of monotonically increasing sequence numbers or unique nonces for each frame, which are included in the data used to calculate the MAC. The receiving SoC will reject any frame that is out of sequence or has a repeated nonce, ensuring the "liveness" of the data stream.

3.3 Trusted Execution Environment (TEE)

While the four pillars secure the data in transit, the processing and signing of the data must occur in a protected environment on the host SoC. SRA designates a Trusted Execution Environment (TEE)—such as Arm TrustZone or Qualcomm's Secure Execution Environment (QSEE/QTEE)—as the "secure chamber" for all sensitive operations. The TEE is a hardware-isolated enclave that runs in parallel to the main operating system (the "normal world") and has exclusive access to protected hardware and memory.

The TEE's responsibilities within the SRA architecture are critical:

1. **Secure Key Management:** The TEE securely stores and manages all cryptographic keys, including the session keys for the encrypted CSI-2 channel and, most importantly, the device's unique private signing key. This private key is generated, stored, and used exclusively within the TEE and is never exposed to the normal world OS or any applications running on it. This design directly mitigates the class of vulnerabilities exploited by security firms like ElcomSoft, who were able to extract private signing keys from the firmware of Canon and Nikon cameras, thereby compromising their entire image authentication systems [4, 9, 10].

2. **Secure Data Processing:** The encrypted and authenticated image stream is passed directly from the ISP to the TEE for decryption and validation. Any necessary image processing, such as format conversion (e.g., RAW to JPEG), occurs within the TEE's isolated memory. This prevents a compromised OS or malicious application from accessing or tampering with the decrypted, unsigned image data.

3. **Cryptographic Signing:** Once the image is processed, the TEE generates a C2PA manifest containing provenance metadata and signs it using the device's private key. The TEE then binds this signed manifest to the final image file before releasing it to the normal world for storage or transmission.

By performing these operations within the TEE, SRA ensures that an unsigned, authentic image never exists in an environment accessible to potentially malicious software. The only asset that ever leaves the secure chamber is the final, cryptographically sealed product.

3.4 Integration with C2PA

For the provenance record to be useful, it must be universally accessible and understandable. SRA leverages the C2PA open standard as its final output format, ensuring that any image signed by an SRA-enabled device is immediately interoperable with a global ecosystem of C2PA-compliant viewers and verification tools [8].

The true power of SRA lies in the unique assertions it embeds within the C2PA manifest. A standard software-based C2PA manifest might record that an image was created by "Device X." An SRA-generated manifest provides a much stronger, verifiable claim. It contains cryptographic proof not just of *what* created the image, but *how* it was created. The manifest generated by the SRA TEE would include specific assertions such as:

- **Hard Binding:** A cryptographic hash (e.g., SHA-256) of the final image pixels, ensuring that any subsequent alteration to the image will be detected [8].

- **Device Identity and Attestation:** A statement signed with the device's unique private key, which can be linked to a certificate from the manufacturer, attesting to the device's authenticity. The manifest can also include a TEE attestation, proving that the signing operation was performed by specific, certified firmware running within the hardware-secure enclave.

- **Secure Pipeline Metadata:** Custom C2PA assertions that explicitly state the security context of the capture. For example: "This image was captured via a secure pipeline," with details like "Encryption: AES-GCM-128," "Authentication: Succeeded," and a unique session or frame counter.

- **Secure Ancillary Data:** Timestamps and geolocation data sourced from trusted hardware components and passed directly to the TEE, rather than being sourced from the potentially compromised normal world OS.

This transforms the C2PA manifest from a simple, self-asserted log file into a verifiable security audit trail. A third-party verifier does not need to trust the device's software; they can cryptographically verify the proof embedded in the manifest. They can see evidence that the image originated from an authenticated sensor, traveled through a protected hardware path, was processed in a secure enclave, and was signed by a key that is cryptographically bound to the device manufacturer. This provides a level of assurance that is an order of magnitude stronger than what software-only solutions can offer.

4 Contribution

To fully appreciate the contribution of the SRA architecture, it is essential to compare it with the MIPI Alliance's Camera Security Framework (CSF), a landmark specification that provides the foundational elements for securing in-device camera pipelines. While SRA is inspired by and builds upon the principles of the MIPI CSF, it serves a distinct and complementary purpose, extending the chain of trust beyond the device to the end consumer of the media.

4.1 MIPI Camera Security Framework (CSF) Overview

The MIPI Camera Security Framework was developed primarily to address the stringent security and safety requirements of the automotive industry, particularly for Advanced Driver-Assistance Systems (ADAS) and autonomous driving applications [14, 16]. In these safety-critical systems, it is paramount that the vehicle's electronic control unit (ECU) can trust that the data it receives from cameras and other sensors is authentic and has not been tampered with [15].

The framework consists of a suite of specifications designed to provide end-to-end, application-level security for data transmitted over MIPI CSI-2 [14, 15]. Its key components include:

- **MIPI Camera Service Extensions (CSE) v2.0:** This specification defines the protocols for embedding security information directly into the CSI-2 data stream. It offers two main methods: Service Extensions Packet (SEP), which adds security headers and footers to each data packet, and Frame-Based Service Extensions Data (FSED), which appends dedicated security packets to an image frame. These protocols carry the Message Authentication Codes (MACs) necessary for integrity protection [15].

- **MIPI Camera Security v1.0:** This specification defines the system-level management of security, leveraging the industry-standard DMTF Security Protocol and Data Model (SPDM) to enable mutual authentication of system components and establish secure sessions [15].

- **MIPI Camera Security Profiles v1.0:** To ensure interoperability, this document defines specific cryptographic suites. The two primary profiles are the "efficiency" ciphersuite, which provides integrity-only protection using AES-CMAC, and the "performance" ciphersuite, which provides both integrity (AES-GMAC) and optional confidentiality (AES-CTR encryption) for systems with dedicated hardware support [11, 15].

The MIPI CSF is a comprehensive and flexible solution for its intended purpose: guaranteeing that the data sink (e.g., an SoC or ECU) within a closed system receives authentic data from the data source (the image sensor).

4.2 Distiction and Project Scope

The fundamental distinction between the MIPI CSF and SRA lies in their scope and ultimate objective. They address different parts of the overall trust problem.

The **MIPI Camera Security Framework's scope is internal**, focused on securing the data pipeline *within* a device or system. Its primary goal is to protect the *process* of data transmission from the sensor to the processor. The trust boundary of the MIPI CSF effectively ends at the SoC. It ensures the SoC can trust its inputs, but it makes no provisions for how a third party can trust the SoC's outputs.

The **Signing Right Away architecture's scope is external**, focused on creating a portable, self-contained, and independently verifiable digital asset. Its primary goal is to protect the final *product*—the image or video file—and imbue it with a verifiable provenance that can be trusted by any third party, anywhere in the world, long after it has left the original device. SRA extends the trust boundary from the SoC to the global information ecosystem.

An effective analogy is to consider the process of minting currency. The MIPI CSF is akin to the armored truck and secure protocols used to transport raw gold (the image data) from the mine (the sensor) to the national mint (the SoC). Its purpose is to prevent theft (eavesdropping) or substitution (tampering) along the way. SRA, by contrast, is the process that happens *inside* the mint. It takes the authenticated gold, verifies its purity, and then stamps it into a serialized, certified coin (the signed C2PA file) whose authenticity can be recognized and trusted by anyone. SRA is the essential "last mile" that transforms a securely transmitted internal data stream into a globally trusted asset.

4.3   Securing Intellectual Property

This distinction in scope makes SRA a powerful tool for securing intellectual property and addressing copyright challenges in ways that a pipeline security protocol alone cannot. By creating a signed, verifiable asset at the point of capture, SRA provides creators with robust, cryptographically-enforceable evidence of authorship and originality.

- **Unforgeable Proof of Authorship:** The digital signature on the C2PA manifest is generated using the device's unique, TEE-protected private key. This signature serves as unforgeable proof that a specific device created the content at a specific time. This can be linked to a registered owner or organization, establishing a clear chain of title from the moment of creation.

- **Secure Timestamping:** The use of a trusted clock, managed within the TEE, to generate the timestamp in the C2PA manifest provides strong, tamper-evident proof of the creation date. This is a critical piece of evidence in establishing priority in copyright disputes.

- **Tamper-Evident Licensing Information:** The C2PA manifest can include assertions for copyright notices, creator information, and machine-readable rights expressions (e.g., Creative Commons licenses). Because this information is part of the cryptographically sealed manifest, it cannot be stripped or altered without invalidating the signature. This ensures that the creator's intended usage rights travel with the content.

This architecture provides a potent defense against a sophisticated form of infringement that could be termed "provenance laundering." In a typical scenario, an infringer might take a copyrighted image, apply minor edits, strip its EXIF metadata, and re-publish it as their own original work. With SRA, this becomes significantly more difficult. The original image would possess an immutable C2PA manifest linking it to the true creator and creation time. The infringer's edited version would either have a broken signature (if they kept the manifest but changed the pixels) or no signature at all (if they stripped the manifest). In a future ecosystem where SRA-signed content is common, an image claiming to be original but lacking a valid provenance record would be inherently suspect. This shifts the burden of proof, making it harder for infringing content to be passed off as legitimate and easier for creators to defend their work.

For maximum resilience, SRA can be used in a hybrid approach with digital watermarking. SRA provides the primary, authoritative, and cryptographically secure proof of origin. A robust digital watermark can be added as a secondary, resilient layer for tracking content that has been screenshotted or transcoded—actions that would break the C2PA signature but might leave a watermark intact. The SRA manifest could even include a signed assertion stating, "A 'Type-X' digital watermark was embedded at the time of capture," linking the two technologies together.

5   Development

The SRA architecture is not merely a theoretical construct; it is the result of a pragmatic development process informed by empirical prototyping and a keen awareness of the evolving technology landscape. The project's evolution from an initial hardware-centric prototype to a mature, ecosystem-aligned framework demonstrates a data-driven approach to solving the complex challenges of content provenance.

5.1 Initial Prototyping and Empirical Findings

The initial phase of the SRA project focused on building a proof-of-concept prototype to validate the core concept of real-time cryptographic signing of a camera feed. This prototype consisted of a Sony IMX219 camera module, an Efinix Trion T20 FPGA for processing, and a Raspberry Pi as the receiving host, using the 'raspiraw' utility on an older version of its operating system. This phase, rather than being a simple implementation, became a crucial research endeavor that yielded invaluable insights that directly shaped the final SRA architecture.

- **Finding 1: The Practical Limits of CSI-2 Metadata.** The initial plan was to transmit the AES-CMAC authentication tag from the FPGA to the Raspberry Pi using the standard 'User-Defined 8-bit' data types (0x30-0x37) within the CSI-2 protocol. However, investigation revealed that the Raspberry Pi's CSI-2 receiver (Unicam) did not support these standard types, instead routing any non-image packets to a generic metadata buffer without clear parsing rules. To overcome this, a clever workaround was devised: the CMAC tag was encapsulated in a packet with a valid but unused image data type identifier (e.g., RGB888 - 0x24) and prefixed with a unique custom header byte (0x0B). This allowed the 'raspiraw' utility's metadata decoding function to be modified to specifically look for this header and correctly parse the CMAC tag. This experience highlighted the real-world challenges of interoperability and the need for a robust, well-defined protocol for security metadata, as later specified in the MIPI CSE standard.

- **Finding 2: The Prohibitive Cost of Real-Time Software Cryptography.** The core of the prototype was a custom AES-CMAC module designed in Verilog for the FPGA. While functional, simulation and testing revealed a significant performance bottleneck. The processing of a single 1920x1232 resolution frame to generate a CMAC tag required approximately ten million clock cycles. To meet the target frame rate of 30 frames per second (fps), a simple calculation revealed the demanding hardware requirement:

$$10,000,000 \frac{\text{cycles}}{\text{frame}} \times 30 \frac{\text{frames}}{\text{second}} = 300,000,000 \frac{\text{cycles}}{\text{second}}$$

This meant the FPGA's clock frequency would need to be at least 300 MHz, and likely higher to account for overhead, exceeding the practical limits of the designed module on the target hardware. This finding was a decisive moment, proving that real-time, high-resolution authenticated encryption is not feasible with soft-core cryptographic implementations on general-purpose FPGAs. It established a clear requirement for dedicated, hardware-accelerated cryptographic engines to make the SRA vision practical.

5.2 Implementation Roadmap

The lessons learned from the initial prototype led to a strategic pivot. The focus shifted from developing custom hardware for every function to a more agile model of implementing SRA firmware on top of existing, commercially available trusted hardware. This led to the creation of a mature, five-stage development plan designed to systematically build and deploy the full SRA architecture. This roadmap, summarized in Table 5.2, demonstrates a logical progression from baseline functionality to full ecosystem integration.

5.3 Aligning with the Broader Ecosystem

The final stage of the roadmap reflects a critical strategic insight: the most effective path to widespread adoption is not to reinvent the wheel, but to align with and build upon the secure hardware capabilities that are already being integrated into commercial SoCs. The initial project plan considered designing custom Application-Specific Integrated Circuits (ASICs) for SRA functions. However, the prototype experience and an analysis of market trends revealed a more efficient path.

Mobile SoC vendors like Qualcomm have already integrated the necessary hardware primitives—such as secure ISPs, hardware crypto accelerators, and robust TEEs—into their platforms, often driven by the need for features like secure biometric authentication [23]. The emergence of the Qualcomm Snapdragon 8 Gen 3 as the first C2PA-compliant mobile platform is a powerful validation of this trend [24].

Furthermore, industry pioneers like Truepic have already demonstrated a similar architecture in practice. Their Foresight system leverages the Qualcomm TEE and the secure hardware pipeline of the Spectra ISP to create hardware-secured, authenticated photos on Snapdragon-powered devices [23]. This serves as a powerful proof-of-concept for the SRA model and demonstrates a clear path to market.

SRA's strategy is therefore to position itself as an open, interoperable reference architecture that can be implemented on any SoC that provides the necessary trusted hardware components. By leveraging existing secure

Stage	Key Objective	Core Technologies	Validation Milestone	Connection to Prototype Findings
1: Baseline	Establish a functional, non-secure camera pipeline.	MIPI CSI-2, FPGA/ISP, Basic image processing (Bayer to RGB, JPEG).	A clear video stream is successfully transmitted from sensor to host.	Re-establishes the baseline system from the prototype but with a forward-looking, production-oriented design.
2: Secure Channel	Implement an authenticated and encrypted camera feed.	Authenticated Encryption with Associated Data (AEAD) cipher (e.g., AES-GCM), mutual authentication protocol (e.g., SPDM).	The host successfully receives, decrypts, and validates the integrity of the encrypted video stream in real time.	Directly addresses the performance bottleneck by specifying the use of hardware-accelerated AEAD ciphers instead of a soft-core implementation.
3: C2PA Integration	Digitally sign the content and bind provenance metadata.	C2PA Rust SDK, digital signatures (ECDSA/RSA), custom C2PA assertions.	A valid, signed C2PA manifest is generated for each captured image, containing secure pipeline metadata.	Implements the "signing" part of SRA, moving beyond the prototype's focus on integrity (CMAC) to full provenance.
4: TEE Hardening	Isolate all sensitive operations within a secure enclave.	Arm TrustZone / Qualcomm TEE, secure key storage, TEE attestation.	The entire process from decryption to signing occurs within the TEE, with private keys never exposed to the normal OS.	Provides the secure environment for key management that was outside the scope of the initial hardware-focused prototype.
5: SoC Integration	Port the SRA framework to commercial mobile SoCs.	Qualcomm Secure Camera APIs, Android HAL, OEM partnerships.	SRA functions as a "Secure Capture Mode" on a commercial-grade smartphone, producing C2PA-compliant assets.	Represents the strategic shift to leveraging existing trusted hardware, making SRA scalable and commercially viable.

Table 2: SRA Phased Development Roadmap

camera APIs and TEE SDKs, SRA can be deployed as a firmware or software solution that "lights up" the latent security capabilities of modern devices. This approach dramatically reduces the cost and time-to-market compared to a custom silicon strategy, and it fosters a competitive, multi-vendor ecosystem rather than a single proprietary solution. This pragmatic pivot, informed by both empirical research and market observation, positions SRA not as a disruptive replacement for existing technologies, but as an essential, value-added layer that is poised for rapid adoption.

6 Conclusion

The digital world is contending with an unprecedented crisis of authenticity. The dual threats of hyper-realistic generative AI and fundamental hardware vulnerabilities have systematically dismantled the traditional foundations of trust in visual media. In this new landscape, incremental solutions are insufficient. A new architectural standard is required—one that is rooted in verifiable hardware security and extends trust from the moment of creation to the point of consumption.

The Signing Right Away (SRA) architecture provides this standard. By establishing a cryptographically secure and authenticated channel from the image sensor to a Trusted Execution Environment, SRA closes the critical hardware vulnerability that allows for the injection of fabricated content. By performing all sensitive processing and signing within a hardware-isolated enclave, it protects the integrity of the provenance process from software-level compromise. Finally, by outputting a C2PA-compliant manifest enriched with verifiable security metadata, it creates a portable, interoperable, and globally trusted digital asset.

SRA does not seek to replace existing standards but to complete them. It builds upon the principles of the MIPI Camera Security Framework, extending its internal pipeline protection to create a verifiable external product. It enhances the C2PA standard, transforming its provenance manifests from self-asserted logs into cryptographically-backed security audit trails. It provides a foundational layer of trust upon which other technologies, such as digital watermarking for copyright tracking, can be built.

The development roadmap for SRA, from an ambitious hardware prototype to a mature, ecosystem-aligned framework, demonstrates a commitment to solving this problem in a pragmatic and scalable manner. By aligning with the capabilities of modern mobile SoCs, SRA charts a clear path to widespread adoption. The implementation of this architecture represents a critical step towards restoring a shared sense of visual reality, fostering a future where every critical image and video—whether used in a courtroom, a news report, or an insurance claim—can be fundamentally trusted at its source.

References

[1] James Betker, Gabriel Goh, Li Jing, Tim Brooks, Jianfeng Wang, Linjie Li, Long Ouyang, Juntang Zhuang, Joyce Lee, Yufei Guo, Waseem Manassra, Prafulla Dhariwal, Casey Chu, Yunxin Jiao, and Aditya Ramesh. Improving image generation with better captions. https://cdn.openai.com/papers/dall-e-3.pdf, 2023. Accessed: 2024-03-26.

[2] Zalán Borsos, Raphaël Marinier, Damien Vincent, Eugene Kharitonov, Olivier Pietquin, Matt Sharifi, Dominik Roblek, Olivier Teboul, David Grangier, Marco Tagliasacchi, and Neil Zeghidour. Audiolm: a language modeling approach to audio generation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 31:1906–1917, 2023.

[3] Tim Brooks, Bill Peebles, Connor Holmes, Will DePue, Yufei Guo, Li Jing, David Schnurr, Joe Taylor, Troy Luhman, Eric Luhman, Clarence Ng, Ricky Wang, and Aditya Ramesh. Video generation models as world simulators. 2024.

[4] Cardinal Peak. ElcomSoft's Hack of Image Authentication. https://www.cardinalpeak.com/blog/elcoms ofts-hack-of-image-authentication, 2011. Accessed: 2024-10-29.

[5] Lucy Chai, David Bau, Ser-Nam Lim, and Phillip Isola. What makes fake images detectable? understanding properties that generalize. In *Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVI*, page 103–120, Berlin, Heidelberg, 2020. Springer-Verlag.

[6] Coalition for Content Provenance and Authenticity (C2PA). C2PA Explainer. https://spec.c2pa.org/sp ecifications/specifications/1.3/explainer/Explainer.html, 2023. Accessed: 2024-10-29.

[7] Coalition for Content Provenance and Authenticity (C2PA). C2PA Specification Version 1.3. https://c2pa .org/specifications/specifications/1.3/index.html, 2023. Accessed: 2024-03-26.

[8] Coalition for Content Provenance and Authenticity (C2PA). C2PA Technical Specification. https://spec .c2pa.org/specifications/specifications/1.3/specs/C2PA_Specification.html, 2023. Accessed: 2024-10-29.

[9] ElcomSoft Co. Ltd. Canon original data security system compromised: Elcomsoft discovers vulnerability. https://www.elcomsoft.com/PR/canon_101130_en.pdf, November 2010. Accessed: 2024-03-26.

[10] ElcomSoft Co. Ltd. Elcomsoft discovers vulnerability in nikon's image authentication system. https://www.elcomsoft.com/PR/nikon_110428_en.pdf, April 2011. Accessed: 2024-03-26.

[11] Andrew Elias and SriDeepti Pisipati. Streamlining camera security validation framework using synopsys mipi cse v2.0 vip. https://www.synopsys.com/blogs/chip-design/streamlining-camera-security-validat ion-framework-using-synopsys-mipi-cse-v20-vip.html, May 2024. Accessed: 2024-10-29.

[12] Geekworm Official Store. Geekworm c792 hdmi to csi-2 converter module with hdmi loop out. https: //www.aliexpress.com/item/1005009392164078.html, March 2024. Accessed: 2025-10-05.

[13] Google DeepMind. Identifying AI-generated content with SynthID. https://deepmind.google/technologi es/synthid/, 2023. Accessed: 2024-03-27.

[14] MIPI Alliance. Mipi alliance releases camera security specifications for flexible end-to-end protection of automotive image sensor data. https://www.mipi.org/press-releases/mipi-releases-camera-security-s pecifications-for-flexible-end-to-end-protection-of-automotive-image-sensor-data, October 2024. Press Release.

[15] MIPI Alliance. MIPI Camera Security Framework (CSE, CCISE, Profiles). https://www.mipi.org/specifi cations/mipi-camera-security, 2024. Accessed: 2024-10-29.

[16] MIPI Alliance. MIPI Camera Security White Paper. https://www.mipi.org/download-the-mipi-camer a-security-white-paper, 2024. Accessed: 2024-10-29.

[17] MIPI Alliance. MIPI Camera Serial Interface 2 (CSI-2). https://www.mipi.org/specifications/csi-2, March 2024. Accessed: 2024-03-26.

[18] Donie O'Sullivan and Jon Passantino. 'verified' twitter accounts share fake image of 'explosion' near pentagon, causing confusion, 2023. Updated 11:35 AM EDT, Tue May 23, 2023.

[19] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. https://arxiv.org/abs/2204.06125, 2022. Accessed: 2024-03-26.

[20] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. Zero-shot text-to-image generation. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 8821–8831. PMLR, 18–24 Jul 2021.

[21] Andreas Rössler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Niessner. Faceforensics++: Learning to detect manipulated facial images. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1–11, 2019.

[22] Toshiba Electronic Devices & Storage Corporation. TC358743XBG, HDMI to MIPI CSI-2 Bridge. https://toshiba.semicon-storage.com/info/docget.jsp?did=14220. Product Datasheet. Accessed: 2024-10-29.

[23] Truepic. Truepic Breakthrough Charts a Path for Restoring Trust in Photos and Videos at Internet Scale. https://www.truepic.com/blog/truepic-breakthrough-charts-a-path-for-restoring-trust-in-photos-and-videos-at-internet-scale, 2020. Accessed: 2024-10-29.

[24] Truepic. Truepic's technology seamlessly embeds the coalition for content provenance and authenticity's (c2pa) content credentials powered by the snapdragon 8 gen 3 platform. https://www.truepic.com/blog/truepic-unveils-watershed-gen-ai-transparency-directly-on-devices-powered-by-snapdragon-mobile-platform, October 2023. Accessed: 2024-10-29.