MULTI-MODAL SENSOR FUSION FOR PHYSICAL PROVENANCE ATTESTATION

A Proof-of-Concept Implementation

Alan Michael Ethington | Independent Researcher | Seattle, Washington, USA | alan@washingtontec.com

ABSTRACT

Current digital provenance standards such as C2PA rely primarily on cryptographic signatures applied to media containers post-capture, leaving them vulnerable to "analog hole" attacks where synthetic content is displayed and re-captured by authenticated devices. We present a proof-of-concept system that correlates multiple physical sensor modalities—quantum shot noise characteristics, inertial measurement unit (IMU) data, and optical flow—to establish evidence of authentic physical capture. Our prototype implements three concurrent verification layers: (1) Poisson noise distribution analysis of raw sensor data, (2) epipolar geometry consistency checking between device motion and visual parallax, and (3) spectral analysis of hand-tremor characteristics to distinguish human operators from mechanical stabilization systems. We demonstrate successful integration of these techniques on commodity Android hardware (Motorola Moto G-Series) and present initial validation results. This work represents an early-stage exploration of multi-modal physical attestation; extensive adversarial testing, cross-device validation, and large-scale deployment studies remain as critical future work.

Keywords: *digital provenance, sensor fusion, hardware security, media authentication, analog hole*

1. INTRODUCTION

The proliferation of generative AI models capable of producing photorealistic synthetic media has created significant challenges for digital evidence authentication. Traditional forensic techniques based on compression artifacts and noise patterns are increasingly ineffective. While cryptographic provenance standards like C2PA provide post-capture integrity guarantees, they do not address fundamental vulnerabilities in the capture pipeline itself.

1.1 The Analog Hole Problem

The "analog hole" refers to the vulnerability where digital content can be displayed on a screen and re-captured through an authenticated device, thereby acquiring a legitimate signature despite synthetic origin. This attack vector bypasses cryptographic protections by operating at the physical layer.

1.2 Research Questions

This work explores:

1. Can multi-modal sensor fusion detect analog hole attacks on commodity hardware?
2. What physical properties distinguish authentic capture from screen recapture?
3. What are the practical engineering challenges in implementing such a system?
4. What fundamental limitations and attack vectors remain unaddressed?

We present initial findings from a working prototype while acknowledging that comprehensive validation remains future work.

2. RELATED WORK

2.1 Cryptographic Provenance

C2PA and similar standards provide tamper-evident signatures using trusted execution environments (TEEs). However, these systems authenticate the signing device rather than the physical capture event.

2.2 Sensor-Based Authentication

Prior work has explored individual sensor modalities for authentication. Lukas et al. demonstrated photo-response non-uniformity (PRNU) for camera fingerprinting. Amerini et al. used IMU data for video authentication. Our approach differs by requiring concurrent multi-modal consistency.

2.3 Deepfake Detection

Traditional deepfake detection focuses on identifying synthetic artifacts in completed media. Our approach

instead verifies physical capture characteristics at acquisition time.

3. SYSTEM DESIGN

3.1 Architecture Overview

Our system implements three concurrent verification layers:

3.1.1 Layer 1: Quantum Shot Noise Analysis

CMOS image sensors exhibit quantum shot noise following a Poisson distribution. We analyze raw Bayer-pattern sensor data before image signal processor (ISP) intervention to compute the Fano factor:

$$F = \sigma^2 / \mu$$

For authentic sensor capture with gain g and read noise σ_{read}:

$$|(\sigma^2 - \sigma^2_{read}) / (g \cdot \mu) - 1| < \tau$$

We use $\tau = 0.07$ based on our test device characteristics. Generative models may inject Gaussian or uniform noise that deviates from Poisson statistics.

Known Limitations:

- Advanced adversaries could simulate Poisson noise
- Temperature affects noise characteristics
- Threshold is device-specific

3.1.2 Layer 2: Epipolar Geometry Consistency

During a required device tilting gesture (~45 degrees), we correlate IMU-reported angular displacement with optical flow analysis. For three-dimensional scenes, optical flow variance should reflect depth discontinuities:

$$\sigma^2_{flow} > 0.12$$

Screen recaptures exhibit near-planar homography with lower variance values.

Known Limitations:

- Curved displays may defeat this check
- Requires user cooperation (gesture)
- Threshold determined empirically from limited testing

3.1.3 Layer 3: Biomechanical Motion Profiling

Human hand tremor exhibits characteristic spectral properties in the 8-12 Hz range. We compute normalized bicoherence to detect quadratic phase coupling:

$$b^2(f_1, f_2) = |E[X(f_1)X(f_2)X^*(f_1+f_2)]|^2 / (E[|X(f_1)X(f_2)|^2]E[|X(f_1+f_2)|^2])$$

Initial testing suggests human motion produces $b^2 < 0.40$, while PID-controlled gimbals exhibit $b^2 > 0.80$.

Known Limitations:

- Excludes users with tremor disorders, mobility aids, or accessibility needs
- Threshold based on limited testing of able-bodied adults
- Sophisticated robots could potentially simulate chaotic motion
- Conflicts with legitimate tripod/mount use cases

3.2 Hardware Integration

Implementation on Android using:

- Camera2 API for RAW sensor access
- Direct sensor-to-TEE data path where supported
- ARM TrustZone for isolated computation

Engineering challenges encountered:

1. Surface locking conflicts between camera HAL and UI
2. Non-standard memory stride requiring custom parsing
3. Hardware stabilization (OIS/EIS) dampening biological signals

3.3 Cryptographic Binding

The system generates a hash chain binding sensor frames and IMU data:

$$H_n = SHA\text{-}256(Frame_n \,||\, IMU_n \,||\, H_{n-1})$$

This chain is signed by the device TEE, binding hardware identity to the capture sequence.

4. PRELIMINARY VALIDATION

4.1 Test Environment

We implemented the system on a single device: Motorola Moto G-Series (Model ZT4222V46R). Testing included:

- Authentic handheld captures in various lighting conditions
- Screen recapture attacks (filming content displayed on an 8K monitor)
- Gimbal-stabilized captures (DJI Osmo Mobile)

Critical limitation: These tests represent initial feasibility validation only. We have not conducted:

- Adversarial red team testing
- Cross-device validation across manufacturers
- Large-scale statistical sampling
- Long-term calibration stability studies

4.2 Observed Results

Shot Noise Analysis:

- Authentic captures showed Fano factor $F \approx 1.02$ (consistent with Poisson distribution)
- Simple synthetic images with Gaussian noise showed $F \approx 1.21$
- **Caveat:** We did not test against adversarial noise models trained on sensor characteristics

Geometric Consistency:

- Authentic 3D scenes: $\sigma^2_{flow} \approx 0.14$
- Flat screen recapture: $\sigma^2_{flow} \approx 0.008$
- Clear separation in this limited test set
- **Caveat:** Did not test curved displays, stereoscopic systems, or light-field displays

Motion Profiling:

- Handheld capture: $b^2 \approx 0.34$
- Gimbal capture: $b^2 \approx 0.82$
- Separation observed in limited samples
- **Caveat:** Did not test sophisticated chaotic motion generators or diverse user populations

4.3 What We Did NOT Test

This is critical to state explicitly:

1. **No adversarial testing:** We did not hire security researchers to attempt defeats
2. **No cross-device validation:** Only tested on one phone model
3. **No sample size determination:** We have not established statistical power
4. **No calibration drift studies:** Unknown how thresholds change over time
5. **No accessibility testing:** Unknown impact on users with disabilities
6. **No sophisticated attack simulation:** Did not test learned noise models, curved displays, or advanced motion synthesis
7. **No false positive/negative rate characterization:** Current numbers are anecdotal

5. LIMITATIONS AND OPEN CHALLENGES

5.1 Fundamental Vulnerabilities

Layer 1 (Shot Noise):

- Adversaries could train GANs on device-specific noise characteristics
- Thermal manipulation could potentially spoof entropy
- Future generative models may include accurate sensor noise simulation

Layer 2 (Epipolar):

- Curved displays likely reduce detection confidence
- Light-field displays could potentially defeat this entirely
- Requires user gesture (UX friction)

Layer 3 (Tremor):

- Excludes legitimate use cases (tripods, accessibility devices)
- Sophisticated robotics could simulate biological chaos
- Raises privacy concerns (biomechanical profiling)
- Potentially discriminatory against users with movement disorders

5.2 Deployment Barriers

Hardware Fragmentation:

- Requires RAW sensor API (not available on all Android devices)
- TEE capabilities vary by manufacturer
- Threshold calibration appears device-specific

Calibration Stability:

- Unknown: How do thresholds drift with sensor aging?
- Unknown: Impact of OS/firmware updates on sensor characteristics
- Unknown: Recalibration frequency required

User Experience:

- Requires deliberate gesture (3+ seconds per capture)
- May fail in low-light conditions
- Blocks legitimate professional use cases

5.3 Unsolved Research Questions

1. **What is the actual cost to defeat this system?** We claim "economic friction" but have not modeled actual adversarial costs
2. **How does this perform across diverse hardware?** Cross-manufacturer validation needed
3. **What are the false positive/negative rates?** Requires large-scale testing
4. **How quickly can adversaries adapt?** Unknown timeline for defeat techniques
5. **What is the accessibility impact?** Requires diverse user population testing

6. FUTURE WORK

To advance this research beyond proof-of-concept:

6.1 Required Validation Studies

1. **Adversarial Red Team:** Partner with security researchers to attempt sophisticated defeats
2. **Cross-Device Testing:** Validate across 20+ smartphone models from different manufacturers
3. **Large-Scale Sampling:** Establish statistical significance with hundreds of samples per category
4. **Longitudinal Calibration Study:** 12+ month deployment tracking threshold drift

5. **Accessibility Research:** Testing with diverse populations including users with disabilities

6.2 Technical Extensions

1. Video capture with temporal consistency
2. Privacy-preserving variants using zero-knowledge proofs
3. Graceful degradation strategies for partial sensor availability
4. Alternative verification paths for excluded populations

6.3 Ecosystem Engagement

1. Standardization discussions with C2PA working groups
2. Engagement with Android AOSP for forensic API development
3. Legal admissibility research (Daubert hearings, expert testimony frameworks)
4. Privacy impact assessment (GDPR compliance for biometric data)

7. DISCUSSION

7.1 What This Work Demonstrates

We have shown that:

- Multi-modal sensor fusion is implementable on commodity hardware
- Initial separation exists between authentic capture and simple attacks in our limited testing
- Integration challenges are solvable with sufficient engineering effort

7.2 What This Work Does NOT Demonstrate

We have NOT shown:

- Robustness against sophisticated adversaries
- Generalization across hardware platforms
- Long-term stability or maintainability
- Acceptable false positive/negative rates at scale
- Compatibility with accessibility requirements
- Economic viability of deployment

7.3 Honest Assessment

This represents **early-stage exploratory research**, not a production-ready solution. The gap between "works on one phone in limited testing" and "deployable security system" is substantial and requires:

- Extensive adversarial testing (6-12 months minimum)
- Multi-institution validation studies
- Standardization and regulatory review
- Accessibility and fairness auditing
- Industry coordination for hardware support

The timeline for any real-world deployment would be measured in years, not months.

7.4 Adversarial Cost Sketch: Practical Barriers and Attack Economics

This section provides a qualitative, order-of-magnitude assessment of the **economic and engineering costs** required for adversaries to defeat the proposed multi-modal physical attestation system. We emphasize that this analysis is **not empirical**, does **not claim completeness**, and is intended solely to contextualize the "economic friction" argument made elsewhere in this work.

A. Attacker Classes Considered

We consider three broad adversary classes, reflecting common threat modeling practice:

1. **Opportunistic Adversary:** Individuals or small groups using readily available software tools, consumer displays, and commodity hardware. Typical use cases include misinformation, impersonation, or low-scale fraud.
2. **Professional Fraud Operations:** Well-resourced criminal groups capable of custom software development, hardware modification, and coordinated attack workflows. Motivated by financial gain.
3. **Well-Resourced / State-Level Actors:** Adversaries with access to specialized hardware fabrication, optical systems, and multidisciplinary engineering teams. Motivated by intelligence, influence, or strategic objectives.

This work is primarily concerned with the first two classes.

B. Layer-by-Layer Adversarial Cost Analysis

Layer 1: Quantum Shot Noise Consistency

Baseline Attack: Inject synthetic noise into generated images to approximate Poisson statistics.

Observed Barrier: While Poisson-distributed noise is straightforward to simulate in isolation, accurately matching **device-specific sensor behavior** (gain, read noise, temperature dependence, pixel-level correlations) requires:

- Access to raw sensor characterization data
- Per-device noise modeling
- Adaptation to thermal and ISO variability

Estimated Cost Impact:

- Opportunistic adversary: Low–moderate (software-only, but brittle)
- Professional fraud group: Moderate (device-specific model training required)
- State-level actor: Low (well within capability)

This layer alone does not provide strong security, but establishes a **minimum entropy requirement** that naive synthetic generation often fails to meet.

Layer 2: Epipolar Geometry and Motion–Vision Consistency

Baseline Attack: Display synthetic content on a high-resolution screen and re-capture during a device motion gesture.

Observed Barrier: Defeating this layer requires presenting **true depth-dependent parallax** synchronized with real-time IMU-reported motion. This is difficult to achieve with:

- Flat displays (planar homography)
- Pre-rendered content without real-time motion coupling

Potential defeat strategies include:

- Curved or multi-plane displays
- Real-time rendered light-field or volumetric displays
- Mechanically synchronized motion rigs

Estimated Cost Impact:

- Opportunistic adversary: High (infeasible)
- Professional fraud group: High (custom optical hardware required)
- State-level actor: Moderate (possible with specialized systems)

This layer imposes **significant physical and synchronization complexity**, particularly in handheld form factors.

Layer 3: Biomechanical Motion Profiling

Baseline Attack: Use mechanical stabilization (gimbal, tripod) to control motion characteristics.

Observed Barrier: PID-controlled stabilization systems exhibit deterministic phase coupling distinct from human tremor. To defeat this layer, an adversary would need to:

- Generate motion with appropriate spectral entropy
- Avoid deterministic control algorithms
- Synchronize motion with visual and IMU signals

Potential approaches include chaotic motion generators or learned motion synthesis systems.

Estimated Cost Impact:

- Opportunistic adversary: Very high (infeasible)
- Professional fraud group: Moderate–high (custom robotics required)
- State-level actor: Low–moderate

This layer is effective against **commodity stabilization attacks**, but does not claim resistance to sophisticated robotic simulation.

C. Combined System Cost Effects

Critically, defeating the full system requires **simultaneous success across all three layers**, under real-time constraints, on a per-device basis:

- Noise statistics must match sensor physics
- Visual parallax must match IMU-reported motion
- Motion characteristics must resemble human biomechanics
- All signals must be temporally synchronized and bound cryptographically

This coupling significantly increases adversarial complexity compared to defeating any single layer in isolation.

D. Summary Assessment

This system does **not** provide cryptographic unforgeability. However, it appears to:

- Block opportunistic and low-effort analog hole attacks
- Impose meaningful engineering and hardware costs on professional fraud operations
- Remain vulnerable to well-resourced, specialized adversaries

We therefore characterize the security benefit as **economic deterrence**, not absolute prevention. Quantifying these costs precisely remains an open research problem and a priority for future adversarial testing.

8. CONCLUSION

We have presented a proof-of-concept system for multi-modal physical attestation of digital media capture. By correlating quantum shot noise, epipolar geometry, and biomechanical motion characteristics, our prototype demonstrates initial feasibility of detecting analog hole attacks on commodity hardware.

However, we emphasize that this work represents the beginning of a research program, not its conclusion. Extensive validation, adversarial testing, cross-platform development, and accessibility research are required before any claims of practical security can be made.

The fundamental challenge remains: as generative AI capabilities improve and adversaries develop adaptive attacks, the gap between authentic and synthetic content continues to narrow. Multi-modal physical verification may provide temporary advantage in an ongoing arms race, but it is not a permanent solution.

We release this work to stimulate research discussion and invite the security community to help identify vulnerabilities, suggest improvements, and conduct independent validation.

ACKNOWLEDGMENTS

This work was conducted independently without institutional funding. We thank the open-source Android development community for Camera2 API documentation and acknowledge the need for peer collaboration to advance this research.

CONFLICT OF INTEREST

The author has commercial interests in Washington Tec, LLC, which may seek to commercialize aspects of this technology.

CODE AND DATA AVAILABILITY

Implementation details are currently proprietary pending patent review. We commit to releasing:

- Sanitized validation datasets upon peer acceptance
- API specifications for independent implementation
- Threshold determination methodology

We invite interested researchers to contact us regarding collaboration opportunities.

NOTE TO REVIEWERS: This manuscript represents work-in-progress. We specifically seek feedback on:

1. Additional attack vectors we should test
2. Suggested methodologies for large-scale validation
3. Accessibility considerations we may have overlooked
4. Related work we should cite
5. Statistical methods appropriate for establishing detection thresholds

REFERENCES

[1] Goodfellow, I., et al. "Generative adversarial networks." NIPS 2014.

[2] Karras, T., et al. "A style-based generator architecture for generative adversarial networks." CVPR 2019.

[3] Verdoliva, L. "Media forensics and deepfakes: an overview." IEEE Journal of Selected Topics in Signal Processing, 2020.

[4] C2PA Specification. "Coalition for Content Provenance and Authenticity Technical Specification v1.0." 2022.

[5] Piva, A. "An overview on image forensics." ISRN Signal Processing, 2013.

[6] Marra, F., et al. "Do GANs leave specific traces?" IEEE WIFS 2019.

[7] Lukas, J., et al. "Digital camera identification from sensor pattern noise." IEEE TIFS 2006.

[8] Amerini, I., et al. "Smartphone fingerprinting combining features of on-board sensors." IEEE TIFS 2017.

[9] Tolosana, R., et al. "Deepfakes and beyond: A survey of face manipulation and fake detection." Information Fusion, 2020.

[10] Rossler, A., et al. "FaceForensics++: Learning to detect manipulated facial images." ICCV 2019.

[11] Janesick, J. R. "Photon transfer." SPIE Press, 2007.

[12] Deng, Y., et al. "Learning to generate realistic noisy images via pixel-level noise-aware adversarial training." NeurIPS 2021.

[13] Deuschl, G., et al. "Consensus statement of the Movement Disorder Society on Tremor." Movement Disorders, 1998.