Post-Quantum Human Signatures in Trusted Hardware

Livy Labs

Short Communication

March 9, 2026

Abstract

This paper presents a hardware-constrained hybrid key-establishment architecture designed to provide cryptographic evidence of human authorization for digital content. To strengthen provenance guarantees, we propose a design that anchors long-term authentication in a Secure Enclave while handling ephemeral session-key derivation through a classical NIST P-256 and post-quantum ML-KEM hybrid exchange. The resulting shared secrets are combined and processed via HKDF to derive fresh session keys with strong key separation and forward-security properties. By separating the static hardware-protected key from the session-key schedule, this architecture preserves compatibility with existing trust anchors, enables a conservative transition to post-quantum security, and tightly binds cryptographic operations to local biometric verification.

1 Introduction

The rapid proliferation of AI-generated content has made it increasingly difficult to distinguish authentic human-created material from synthetic digital outputs. As traditional signals of authenticity become less reliable, there is a growing need for mechanisms that can provide stronger evidence of provenance, authorship, and human authorization.

Trusted hardware offers one practical foundation for this goal. A Trusted Execution Environment (TEE) provides a hardware-isolated environment for security-sensitive components, separated from the regular operating system [1]. In Apple platforms, this role is realized by the Secure Enclave, which provides hardware-backed protection for long-term key material and local authorization policies [2]. Under the assumption that this subsystem remains uncompromised, it can serve as a trust anchor for cryptographic operations tied to human authorization.

At the same time, trusted hardware does not eliminate the need for cryptographic agility. In particular, post-quantum migration is necessary because, as NIST notes, encrypted data is already exposed to the "harvest now, decrypt later" threat, in which adversaries collect protected data today in the hope of decrypting it once cryptographically relevant quantum computers become available [3]. Motivated by this, we consider a hybrid key-establishment architecture in which long-term authentication remains anchored in secure hardware, while ephemeral session-key derivation evolves more flexibly to incorporate post-quantum protection.

Accordingly, the proposed construction combines a mature classical primitive with NIST-standardized ML-KEM so that security does not rely on a single cryptographic assumption during the transition to post-quantum systems [4, 5]. More broadly, the purpose of this architecture is not to establish the truth of arbitrary digital content, but to strengthen provenance and authenticity guarantees by binding cryptographic operations to local biometric authorization and a hardware-protected execution environment.

1.1 Hybrid Key Exchange

In secure hardware-constrained environments, cryptographic agility is often limited by legacy or fixed-function implementations that continue to rely on classical primitives such as the NIST P-256 elliptic curve for discrete-logarithm-based key establishment [6, 7]. To enable a conservative transition toward post-quantum security without significantly increasing implementation complexity, a minimal hybrid key-establishment design combines two components: a classical P-256 key-agreement mechanism and a post-quantum Module-Lattice-Based Key-Encapsulation Mechanism (ML-KEM) [5, 7]. Formally, the hybrid construction is defined as

$$\hat{K} = (K_1, K_2), \tag{1}$$

where K_1 denotes the classical P-256 key-agreement component and K_2 denotes the post-quantum ML-KEM component.

After exchanging the corresponding public values, the communicating parties independently derive the shared secrets Z_1 and Z_2. These values are then combined using a concatenation-based combiner:

$$\hat{Z} = Z_1 \parallel Z_2. \tag{2}$$

This construction is particularly attractive in constrained environments because it introduces minimal protocol overhead while preserving hybrid robustness. Formal analyses show that the establishment of hybrid keys based on concatenation can ensure that the final derived key remains secure as long as at least one component of the system remains secure [4, 8, 9].

Because the concatenated value $\hat{Z}$ is not guaranteed to be uniformly distributed, it should not be used directly as session keying material. Instead, it is processed using the HMAC-based Key Derivation Function (HKDF), introduced by Krawczyk and standardized in Request for Comments (RFC) 5869 [10, 11]. In this construction, HKDF takes $\hat{Z}$ as the input secret, optionally incorporates a salt r, and derives an output key of length l bits according to

$$\mathsf{PRK} = \mathsf{HKDF\text{-}Extract}(r, \hat{Z}), \tag{3}$$

$$K_{\text{out}} = \mathsf{HKDF\text{-}Expand}(\mathsf{PRK}, \mathit{info}, l). \tag{4}$$

This extraction-and-expansion process yields cryptographically strong keying material suitable for encryption and authentication [10, 11]. Consequently, the resulting session key provides robust transitional security against both classical and quantum adversaries, provided that at least one of the two key-establishment components remains secure [4, 8, 9].

1.2 Formal PRF View of Key Derivation

The security of the KDF can be expressed in terms of a pseudorandom function (PRF). Let

$$f : \mathcal{K} \times \mathcal{D} \to \mathcal{R} \tag{5}$$

be a keyed function family. For an efficient adversary A, the PRF distinguishing advantage is defined as

$$\mathrm{Adv}^{\mathrm{prf}}_{A,f}(\kappa) = \left| \Pr_{K \leftarrow \mathcal{K}} \left[A^{f_K}(1^\kappa) = 1 \right] - \Pr_{g \leftarrow \mathrm{Func}(\mathcal{D},\mathcal{R})} \left[A^{g}(1^\kappa) = 1 \right] \right|. \tag{6}$$

The PRF property requires

$$\mathrm{Adv}^{\mathrm{prf}}_{A,f}(\kappa) \leq \mathrm{negl}(\kappa), \tag{7}$$

meaning that no polynomial-time adversary can distinguish the real keyed function from a random function except with negligible advantage. Under this perspective, HKDF extracts entropy from $\hat{Z}$ and expands it into traffic secrets that are computationally indistinguishable from random under standard assumptions on HMAC and the underlying hash function [10, 11].

2 Methodology

To address this problem, we propose a system designed to provide stronger authenticity guarantees in a setting where the gap between real-world evidence and AI-generated content is increasingly significant. More specifically, we aim to attach authenticity properties to user-generated content so that one can obtain evidence that the content originated from a human rather than from a synthetic process. Although several approaches are possible, we take as a starting point a technology that is already widely deployed. In particular, we consider a setting in which a Secure Enclave uses facial information locally to authorize cryptographic operations. Under the trust assumption that this hardware-protected subsystem remains uncompromised, this model motivates the architecture shown in Figure 1.

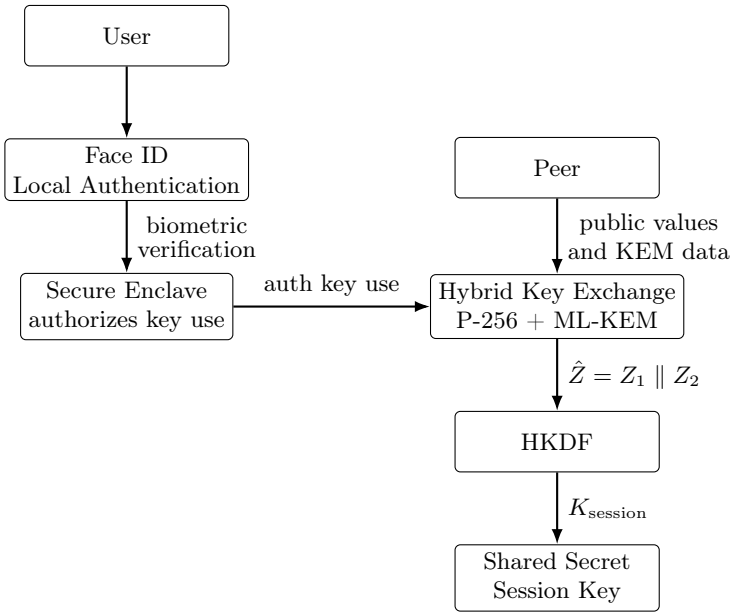

Figure 1: Hybrid key exchange establishment

2.1 Key Refreshing

To reduce computational cost while preserving session-key hygiene, the proposed architecture separates long-term authentication from ephemeral key establishment. A Secure Enclave–resident NIST P-256 key is used only to authenticate the initial handshake, while session secrets are derived from a fresh hybrid exchange rather than from the long-term key itself [2, 6, 12, 13]. This is consistent with modern authenticated key exchange, where signatures authenticate the transcript and forward secrecy is obtained from fresh ephemeral key material [13]. After authentication, the peers perform an initial post-quantum hybrid exchange, combine the resulting shared secret material $\hat{Z}$, and process it through HKDF to derive both traffic keying material and a distinct resumption secret [10, 11, 13].

For later phases, fresh traffic secrets are derived by reinjecting maintained secret state into the KDF, following the TLS 1.3 key schedule and `KeyUpdate` mechanism [10, 13]. If forward secrecy is required across resumed connections, the resumption secret should be

combined with a fresh ephemeral exchange, analogous to PSK-(EC)DHE in TLS 1.3, since PSK-only resumption does not provide the same guarantee [13]. This staged derivation strategy improves key freshness, enforces separation between epochs, and limits exposure under any single traffic secret, while relying on enclave hardware designed to protect sensitive operations and reduce timing and side-channel leakage [2, 12].

2.2 Cryptographic Ownership and Provenance

Within the current scope, the derived key material can also be used to associate ownership and provenance metadata with digital content. In particular, this work is intended to integrate with *zkMedia*, a project aimed at proving and tracing modifications applied to media objects. While the immediate focus is on images, the same approach can be extended to videos and other data types for which each transformation step can be recorded and verified. In that broader setting, the system provides traceability across the full processing pipeline by linking successive operations to authenticated cryptographic evidence. Although several mechanisms can be used to record downstream transformations, the principal problem addressed in this work is earlier in the chain: guaranteeing the authenticity of the input itself. The architecture proposed here is designed to provide that initial trust anchor, upon which stronger end-to-end provenance and ownership guarantees can later be built.

3 Conclusion

As the proliferation of AI-generated media challenges traditional signals of authenticity, establishing verifiable human provenance has become a critical security requirement. This paper presented a hardware-anchored hybrid key-establishment architecture that directly addresses this gap by binding cryptographic operations to local biometric verification within a Secure Enclave, providing a reliable trust anchor at the exact moment of content creation.

The proposed design successfully balances immediate hardware constraints with long-term cryptographic agility. By decoupling the static authentication key from the ephemeral session-key schedule via HKDF, the architecture ensures strong key separation and forward secrecy. Simultaneously, integrating classical NIST P-256 with post-quantum ML-KEM guarantees robust transitional resilience against both current cryptanalytic capabilities and future quantum threats.

Ultimately, this architecture is not just a theoretical key exchange, but a foundational tool for media provenance. By establishing cryptographically verifiable evidence of human origin, it solves the fundamental "input problem" for downstream zero-knowledge (ZK) pipelines and other frameworks designed to prove media transformations. Guaranteeing the authenticity of the source material before any processing occurs ensures that digital content retains a durable, mathematically verifiable chain of trust, offering a practical defense against the unchecked spread of synthetic media.

References

[1] H. Tschofenig, D. Wheeler *et al.*, "Trusted execution environment provisioning (teep) architecture," RFC 9397, 2023.

[2] Apple Inc., "Secure enclave," Apple Platform Security, Dec. 2024. [Online]. Available: https://support.apple.com/en-sg/guide/security/sec59b0b31ff/web

[3] National Institute of Standards and Technology, "Transition to post-quantum cryptography standards," National Institute of Standards and Technology, Tech. Rep. NIST IR 8547 (Initial Public Draft), Nov. 2024.

[4] N. Bindel, J. Brendel, M. Fischlin, B. Gonçalves, and D. Stebila, "Hybrid key encapsulation mechanisms and authenticated key exchange," in *Post-Quantum Cryptography (PQCrypto 2019)*, 2019, pp. 206–226.

[5] National Institute of Standards and Technology, "Module-lattice-based key-encapsulation mechanism standard," National Institute of Standards and Technology, Tech. Rep. FIPS 203, Aug. 2024.

[6] L. Chen, D. Moody, A. Regenscheid, A. Robinson, and K. Randall, "Recommendations for discrete logarithm-based cryptography: Elliptic curve domain parameters," National Institute of Standards and Technology, Tech. Rep. NIST SP 800-186, Feb. 2023.

[7] E. Barker, L. Chen, A. Roginsky, A. Vassilev, and R. Davis, "Recommendation for pair-wise key-establishment schemes using discrete logarithm cryptography," National Institute of Standards and Technology, Tech. Rep. NIST SP 800-56A Rev. 3, Apr. 2018.

[8] M. Campagna and A. Petcher, "Security of hybrid key encapsulation," Cryptology ePrint Archive, Paper 2020/1364, 2020. [Online]. Available: https://eprint.iacr.org/2020/1364

[9] A. Petcher and M. Campagna, "Security of hybrid key establishment using concatenation," Cryptology ePrint Archive, Paper 2023/972, 2023. [Online]. Available: https://eprint.iacr.org/2023/972

[10] H. Krawczyk and P. Eronen, "Hmac-based extract-and-expand key derivation function (hkdf)," RFC Editor, Tech. Rep. RFC 5869, May 2010.

[11] H. Krawczyk, "Cryptographic extraction and key derivation: The hkdf scheme," Cryptology ePrint Archive, Paper 2010/264, 2010. [Online]. Available: https://eprint.iacr.org/2010/264

[12] Apple Inc., "Uses for optic id, face id, and touch id," Apple Platform Security, Dec. 2024. [Online]. Available: https://support.apple.com/en-sg/guide/security/sec067eb0c9e/web

[13] E. Rescorla, "The transport layer security (tls) protocol version 1.3," RFC Editor, Tech. Rep. RFC 8446, Aug. 2018.