



Original Article

Content Provenance Infrastructure: Cryptographic and Behavioural Approaches to Authenticity

Bharath Kandati

Independent Researcher, USA.

Received On: 12/06/2026

Revised On: 09/07/2026

Accepted On: 19/07/2026

Published On: 30/07/2026

Abstract - A platform must decide whether content is authentic when nothing arriving at its front door carries a verifiable history and no upstream party attests to anything. Cryptographic provenance answers that question where it exists [1], but that is a target state. No cryptographic provenance layer existed in the music supply chain this account draws on, so the cryptographic side here is analytical, from the literature and from [1], and the behavioural side is operating practice. The composition that was operated is internal to the behavioural layer and is the first contribution: artefact-level detection is gamed by adversarial post-processing and covered by behavioural context; behavioural context is gamed by patience and covered by identity anchoring; identity is gamed by theft and covered by decision history and adversarial appeal, under the rule that no two signals share a failure mode and no signal's confidence overrides another's alarm. Second, behavioural reconstruction is treated as a provenance layer in its own right, verifying by convergence rather than signature. Third, it sets out an operating policy for content carrying no provenance at all, which is the whole environment, not an edge case: absence triggers measurement rather than trust or suspicion, provenance accretes retroactively from observed behaviour, and claims the platform did not generate are priced as testimony rather than believed. Composition with a cryptographic layer is offered as a proposal, not a deployment. The account is a single-site retrospective practitioner report from a large-scale music streaming platform with, by the author's account, more than 250 million tracks. No controlled evaluation and no operational metrics are reported.

Keywords - Content Provenance, Behavioural Signals, Content Credentials, Synthetic Media Detection, Identity Anchoring, Unprovenanced Content.

1. Introduction

The cost of producing convincing content has collapsed. The cost of proving where content came from has not moved. Generation tools now produce music, voices, and images that pass casual inspection in seconds at near-zero marginal cost, while the infrastructure for answering the question "where did this come from" is substantially what it was a decade ago: metadata that anyone can write, distributor attestations that are self-reported, and file formats that carry no verifiable history. Creation became industrialised while

attribution stayed artisanal. That asymmetry, rather than the bad faith of any particular actor, is what makes provenance unknowable at the moment a platform has to decide something.

In music the supply chain compounds the asymmetry. A track does not arrive at a platform from an artist. It arrives through a chain: creator, possibly a generation tool, an aggregator, a distributor, then the platform. Every hop is a place where provenance information is dropped, flattened, or fabricated, and no hop is contractually obliged to preserve what it received. By the time content reaches the decision point, its history is whatever the metadata says it is, and metadata is testimony, not evidence.

Four things break when provenance is unknown. Attribution breaks first, and money follows it, because royalty systems assume it is known who made the thing; when provenance is unknowable, payment flows to whoever claims the work most convincingly, which advantages fraud operating at industrial scale over legitimate creators operating one track at a time. Impersonation compounds the harm, since a synthetic voice clone does not merely take a stream's royalty, it spends a real artist's reputation without permission. The economics of counterfeited attention in streaming, including spam, click fraud, and fabricated artists, have been documented as a structural feature of the streaming economy rather than an anomaly within it [13].

Second, inference-based detection is forced to exist at all. Detection is the compensating control for missing provenance infrastructure. Audio is analysed to guess at origins because the supply chain cannot report them. Detection is probabilistic, adversarial, and expensive; provenance done properly is a lookup. An inference system operating across hundreds of millions of tracks exists to answer a question that verifiable provenance would have made trivial for most of that content.

Third, trust burdens invert. Without provenance, legitimate creators bear the cost of proving they are real, through verification processes, appeals, and disputed flags, while bad actors exploit the default of anonymity. The honest participants pay the tax and the dishonest ones collect it.

Fourth, the ecosystem's epistemics degrade. Once listeners know that synthetic content is present and unlabelled, suspicion does not stay contained to the synthetic content. It discounts everything. Unknown provenance does not merely misattribute individual items; it cheapens authenticity as a category.

The one-line version is that when provenance is unknown, every downstream system, including payment, recommendation, enforcement, and trust, has to substitute inference for knowledge, and inference at adversarial scale is the most expensive way to know anything.

One thing has to be stated before the argument begins, because it bounds what this paper is entitled to claim. No cryptographic provenance layer existed in the supply chain

this account draws on. Watermarks and signed attestations either were not present in the music supply chain or arrived too unreliably to build on, and there were no cooperative provenance partners, because none existed to have. Everything said here about the behavioural layer is therefore an account of operating practice. Everything said about the cryptographic layer, including its deployment failure modes in Section 4 and its composition with the behavioural layer in Section 6.2, is analytical: it is drawn from the standards and research literature and from the author's own prior published architecture [1], not from a system the author operated. The composition that was operated is the one internal to the behavioural layer, set out in Section 6.1, and it is the stronger claim of the two. Section 1.2 states the evidentiary basis in full.

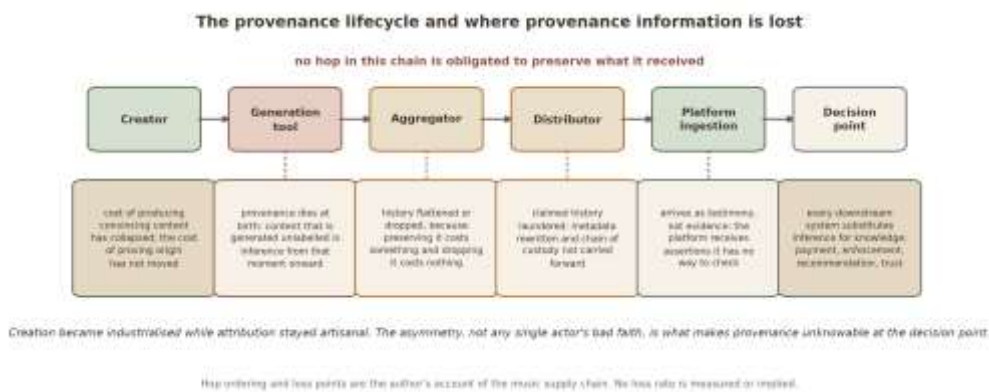


Figure 1. Provenance Lifecycle and Information Loss across the Content Supply Chain

Figure 1. The provenance lifecycle and where provenance information is lost. Hop ordering and loss points reflect the author's account of the music supply chain; no loss rate is measured or implied.

1.1. Contribution Boundary against the Author's Prior Published Architecture

This paper is the second of the author's papers on content provenance, and the boundary between them is drawn explicitly to avoid restating prior work.

The prior paper, "From Detection to Provenance: A Deterministic Architecture for Authenticating AI Generated Content" [1], argues that long-term authenticity cannot depend on probabilistic inference alone and proposes a deterministic cryptographic provenance framework, the Authentication Architecture Layer, which separates generation-time signing, metadata construction, distribution, and independent verification into interoperable layers, establishes authenticity at the point of content creation, and formalises the corresponding system assumptions, threat models, and research agenda. That paper owns the deterministic cryptographic architecture. This paper does not re-present it, does not restate generation-time signing as a novel proposal, and does not claim the layered separation of cryptographic responsibilities as its own contribution. Where the cryptographic layer appears here, it appears as a component whose specification is taken as given from [1]

and is discussed only in terms of how it fails in deployment and how another layer covers those failures.

What this paper adds is the part the prior architecture does not address, because the prior architecture is a description of the target state and this paper is a description of the present one. Three contributions are new.

First, the paper specifies the composition that was operated, which is orthogonal and internal to the behavioural layer: artefact-level detection covered by behavioural context, behavioural context covered by identity anchoring, and identity covered by decision history and adversarial appeal, under the design rule that no two signals may share a failure mode and no signal's confidence may override another's alarm. Second, the behavioural layer is treated as a provenance mechanism in its own right, with its own recording model, its own verification semantics based on convergence rather than signature checking, and its own enumerated failure modes, rather than as the reactive detection the prior paper argues against depending on. Third, and given the most weight here, the paper specifies what to do about content that carries no provenance at all, which is effectively everything arriving at the platform this account draws on, together with a validation posture for provenance claims the platform did not generate. Neither of those problems is a cryptographic problem, neither disappears if the cryptographic layer arrives in full, and neither is treated

in [1]. The composition of the behavioural layer with a cryptographic one is set out separately, in Section 6.2, as a proposal that follows from the same design rule rather than as a result, since no cryptographic layer was available to compose with.

The remainder of the paper proceeds as follows. Section 1.2 states the method and the evidentiary basis. Section 2 places the work in the literature and against the author's prior contributions. Section 3 examines why provenance fails in practice rather than in principle. Sections 4 and 5 enumerate the failure modes of the cryptographic and behavioural layers respectively. Section 6 specifies the composition that was deployed and, separately, the one that is proposed. Section 7 addresses the unsigned majority. Section 8 addresses third-party claims. Section 9 states what has to change across the industry. Section 10 states the limitations, which are substantial and are not minimised.

1.2. Method and Evidentiary Basis

This paper is a single-site retrospective practitioner account, and it is worth stating that as a design rather than leaving it to be inferred, because the design determines what the paper can and cannot support.

Site, domain, and vantage. One platform, one content domain. The account describes an authenticity determination system operated at a large-scale music streaming platform, and the author's vantage is that of the practitioner who led the team that built and ran it. Music properties shaped every conclusion here, in particular a supply chain that concentrates through a relatively small number of distribution pathways and an identity layer, artist verification, that other domains may lack.

How the account was elicited. It was reconstructed from operating experience through a structured retrospective questionnaire, not from extracted logs, records, or a case sample. No observation window is reported and none should be inferred from the retroactive scoring of historical content mentioned in Section 7.4.

What comes from where. The behavioural layer in Section 5, the composition in Section 6.1, the routing rule in Section 6.3, the adoption account in Section 6.4, the unprovenanced-content policy in Section 7, and the claim-validation posture in Section 8 are operating practice. Section 4, the comparison in Table 1's cryptographic column, and the proposed composition in Section 6.2 are analytical: they draw on the standards literature [3], [4], on signing and revocation operations [15], [16], and on the author's prior published architecture [1]. The detection, watermarking, and fingerprinting premises in Sections 2.2, 2.3, 2.4, and 5.3 are literature. Section 9 is the author's opinion, argued from that experience and labelled as opinion.

What is not here. No controlled evaluation. No baseline, no ablation of the composition rule, and no comparison against an alternative architecture. No operational data was released; the author declined to share it. Accordingly there is

no coverage percentage, precision, recall, false positive rate, appeal overturn rate, or human review ratio anywhere in this paper, its tables, or its figures, and the scale statements that do appear are the author's account rather than measurements.

What would count as evidence against the argument. The composition rule claims that orthogonally chosen signals do not fail together. The observation that would falsify it is an adversary who defeated two of the three signal classes in Section 6.1 with one technique, for instance a laundering method that made both the artefact and the surrounding behaviour look ordinary at industrial volume. No such case is reported here, and the record available to this paper does not establish whether one occurred, which is a limitation of the account and not a claim that none exists.

2. Related Work and Prior Contribution

2.1. The Author's Prior Work

Two of the author's prior papers bear on this one. The provenance architecture paper [1] is the direct antecedent and its boundary is drawn in Section 1.1. A companion paper on the societal, ethical, and policy implications of AI-generated content detection [2] argues the case for why detection capability matters beyond the platforms that operate it, and this paper's Section 9 extends that argument in a specific direction, namely that the binding constraints are incentives and liability rather than technology.

One shift in the author's own position is named here rather than left for a reader to find. [1] argues that long-term authenticity cannot depend on probabilistic inference alone. This paper treats behavioural inference as permanent. The two positions are compatible, and the second follows from the first: inference is an insufficient substrate for authenticity and a permanent component of any real system at the same time, because cryptographic coverage can never be retroactive, so whatever it does not reach stays with inference indefinitely.

2.2. Provenance Standards and Content Credentials

The standards work on content provenance is mature enough to deploy. The Coalition for Content Provenance and Authenticity technical specification defines how assertions about an asset's origin and edit history are bound to the asset, carried with it, and verified downstream [3]. Recent systems work demonstrates that the standard can be carried into demanding delivery contexts, including live video, when it is paired with hardware-backed key protection at the point of capture [4]. Both results support this paper's position in Section 9 that provenance at scale is not blocked on cryptography.

The standard also has an answer to the stripping problem, and the argument in this paper has to meet it rather than assume it away, because stripping is the objection Section 4 presses hardest. Alongside the hard cryptographic binding, the specification defines a soft binding as a content identifier that is either a perceptual fingerprint or an invisible watermark, and defines a durable Content Credential as one for which soft bindings exist that allow it to be discovered in

a manifest repository [3]. The intent is precisely the case this paper raises: when a manifest is removed from an asset but a copy of that manifest remains in a provenance store elsewhere, a soft binding lets the content be matched even though the bits differ, and the manifest recovered. That is a real answer and it narrows the objection considerably. It does not close it in the environment described here, for three reasons. Recovery presupposes that a manifest was created and deposited, so it does nothing at all for content that was never signed, which is the dominant case in this account. The soft bindings are themselves watermarks and fingerprints, so they inherit the robustness limits of those mechanisms under deliberate post-processing rather than escaping them, which is the subject of Section 2.4. And a repository lookup is a per-item external dependency at ingestion, which relocates the verification-cost problem of Section 4 rather than removing it. Durable credentials change the shape of the stripping failure mode; they do not retire it.

Music also has identification and delivery infrastructure already, and it is worth separating that from provenance, because the two are routinely conflated. Audio fingerprinting matches a recording against a reference database from the signal itself, robustly across encoding and channel changes, and it is deployed industrially for content identification and rights matching [12]. It answers which recording this is, not where the recording came from. A fingerprint match inherits whatever provenance the reference entry already carried, and newly generated content, being unregistered, matches nothing. The industry's identifier and delivery standards, ISRC for recordings and DDEX delivery messages for the metadata that accompanies them, are the containers that in principle could carry a claimed history through the supply chain, and they are also the containers Section 3 describes being flattened, because they record what a submitter asserts and no hop is obliged to forward what it received. Naming them makes the transit problem concrete: the container exists, and the container is not the binding constraint.

What the standards literature does not settle, and does not claim to settle, is what a platform should do when no credential of any kind is present, which is the case for the entirety of a historical catalogue and for any content produced by a tool that declines to participate.

2.3. Detection, and why it Degrades

The detection literature is where the limits of the inference approach are documented most honestly. Surveys of media forensics and synthetic media detection describe a field in which detectors are trained against the generators that exist and are evaluated against the manipulations that have already been catalogued [5]. Detectors of generated imagery have been shown to be effective and simultaneously fragile, with the strongest early result explicitly hedged as holding "for now" [6], and the arrival of a new generator family has repeatedly required the detection problem to be re-approached rather than merely retrained [7]. In audio, the spoofing and deepfake speech challenges have documented both real progress and the persistent gap between laboratory and in-the-wild conditions [8], and system-level analysis of

cross-domain generalisation failure in audio deepfake detection confirms that performance on held-out domains is the binding constraint rather than performance on the training distribution [9]. This is the empirical basis for the claim, made in [1] and retained here, that detection alone cannot be the long-term substrate of authenticity. It is also the empirical basis for a claim [1] does not make: that no behavioural or inference-based system should report an absolute accuracy figure, because there is no oracle for the ground truth it would be measured against.

2.4. Watermarking, Stripping, and Evasion

Watermarking sits between the cryptographic and behavioural approaches and inherits problems from both. Removal attacks against generative-model watermarking have been demonstrated directly in the image domain [10]. Because the domain of this account is audio, that result on its own would not carry the argument, and it does not have to: dedicated benchmarking of audio watermarking robustness against a broad attack set, including desynchronisation and generative-model attacks, now exists in the same modality [11]. Together they establish that a watermark's presence is evidence and its absence is not, which is the same asymmetry that governs signed metadata, and they establish it without generalising from images to music. This matters for the durable-credential mechanism in Section 2.2 as well, since a soft binding is a watermark or a fingerprint and is therefore subject to the same attack surface. The broader adversarial machine learning literature establishes that any inference-based decision boundary can be approached deliberately, and that evaluating robustness requires assuming an adversary who optimises against the specific defence deployed [14]. Section 5 treats this as a design constraint rather than a research problem: if any single signal class can be defeated by an adversary who knows about it, then the architecture, not the signal, has to carry the resistance.

2.5. Signing Infrastructure, Key Management, and Trust Chains

The operational literature on signing infrastructure is directly relevant to Section 4 and is largely absent from provenance discussions. Work on making software signing broadly deployable establishes that the hard part of a signing ecosystem is identity binding and key management rather than signature computation [15], and measurement work on certificate revocation in the web's public key infrastructure establishes that revocation is the part of a deployed trust chain that most reliably fails to work as specified [16]. Provenance systems will inherit both problems. A generation tool's signing key will eventually leak, and a trust model that cannot revoke without orphaning legitimately signed history will be unable to respond when it does.

2.6. Behavioural and Identity Signals

The behavioural layer described in Section 5 has an established lineage in abuse detection rather than in provenance. The Sybil attack formalises why identity cannot be assumed to be scarce in an open system without a mechanism that makes identity costly [17], which is the

theoretical basis for treating artist verification as an anchor whose value derives from the expense of accumulating it rather than from any cryptographic unforgeability. Work on uncovering large groups of coordinated malicious accounts through the correlation of their actions over time establishes the pattern this paper generalises to provenance: individual items may be indistinguishable while the campaign that produced them is not [18]. The contribution here is to reposition these techniques. In the abuse literature they detect coordinated behaviour. In this paper they reconstruct probable provenance, which is a different question answered with similar evidence.

3. Why Provenance Fails in Practice

Provenance does not fail because the mechanisms are unavailable. It fails at four points that are all economic or contractual rather than technical, and Figure 1 locates the first two of them in the chain.

It dies at birth. The moment content is generated unlabelled, everything downstream is inference. Voluntary attestation by generation tools is the current state, and voluntary means adversarially absent, because the tools chosen for fraud are precisely the tools that will not participate. The generation tool's customer is the uploader, not the ecosystem, so the tool's incentive to label is weak wherever labelling reduces the value of its output to the person paying for it.

It is destroyed in transit. This is the part least discussed and most consequential. Provenance does not merely fail to be created; it gets stripped on the way. Every hop, whether aggregator, distributor, or transcoder, flattens or drops history, because preserving it costs something and dropping it costs nothing. The industry's identifier and delivery formats already provide somewhere for that history to sit, as Section 2.2 notes, so what is missing is not a container but a consequence for emptying one. Section 9 names the lever.

Absence is free. Today unprovenanced content enjoys the same defaults as attested content. Attestation is therefore pure cost with no return, which is a complete explanation of why adoption stalls without needing to attribute bad intent to anyone.

Verification is assumed rather than budgeted. A signature that cannot be checked at ingestion speed, across hundreds of millions of items, is a research contribution rather than infrastructure. The cost of verification at scale is routinely treated as an implementation detail in provenance proposals, including in the target-state architecture of [1], and it is not one.

The consequence of these four failures is that the platform, which is the last actor in the chain and the only one that must decide, receives assertions rather than evidence. It cannot verify them, cannot compel their production, and cannot decline to decide. Everything in the following sections follows from that position.

4. The Cryptographic Layer and Its Failure Modes

The cryptographic layer is specified in [1] and is not re-specified here. What matters for this paper is how it behaves when it exists in a real supply chain, because its failure modes determine what the second layer has to cover.

Cryptographic provenance has one enormous property: it verifies by checking a signature, which means it collapses the authenticity question for a covered item from an inference problem to a lookup. It also produces answers that organisations accept without argument, which, as Section 6.4 discusses, turns out to matter as much as correctness. Five failure modes qualify that property.

Absent at source. Coverage depends entirely on the cooperation of the generating tool. Where signing is voluntary it will be absent exactly where it would be most useful.

Stripped in transit. A signature that arrives detached, or does not arrive, is indistinguishable at the receiving end from content that was never signed, unless the receiver knows independently that this content should have carried one. The specification's durable-credential mechanism is aimed squarely at this failure mode and partially answers it, by allowing a stripped manifest to be recovered from a repository through a soft binding [3], with the three limits set out in Section 2.2: nothing to recover for content never signed, soft bindings subject to the same attacks as any watermark or fingerprint [10], [11], and a per-item external lookup at ingestion.

Forged or misissued credentials. A cryptographic check is a single point of decision, which is its efficiency and its exposure. A credential that verifies is a complete answer only where its signer chains to a trust list the verifier recognises, and trust-list governance is itself a contested part of the deployed design rather than a settled one. Within that boundary the adversary's task reduces to obtaining one credential, whether by compromise, by misissuance, or by finding a recognised issuer with weak identity binding [15].

Key compromise and revocation. Signing keys will be extracted and abused. The trust model therefore needs revocation and rotation that does not retroactively orphan large volumes of legitimately signed items, and the measured behaviour of revocation in the most heavily deployed trust chain in existence is not encouraging [16]. This is unglamorous operational work, which is a substantial part of why it is unfunded.

Verification cost and deployment burden. Verification has to run at ingestion speed at catalogue scale, and the deployments that carry provenance into demanding contexts do so by adding hardware trust anchors at the point of capture [4], which raises rather than lowers the barrier for the long tail of creators.

Two of these failure modes, absence and stripping, share a structural consequence worth stating precisely, because Section 7 depends on it. A cryptographic signal is asymmetric evidence. Its presence, verified, is strong evidence. Its absence is not evidence of anything at all, as long as absence is the norm. Any system that treats missing attestation as suspicion, before attestation is universal, is not enforcing provenance; it is penalising the people least equipped to produce it.

5. The Behavioural Layer and Its Failure Modes

The behavioural layer is what a platform builds when cryptographic provenance has failed upstream of it. It is worth being precise about what it is, because it is not the cryptographic architecture with the cryptography removed. It records different things, verifies by a different mechanism, and fails in different ways.

5.1. What is Recorded, by whom

The behavioural layer records signals rather than attestations. Four classes were used in the system this paper draws on. Audio-level characteristics that correlate with generation methods. Contextual behaviour, including upload timing and volume patterns, metadata consistency, catalogue structure, and distributor pathways. Identity signals derived from artist verification, which anchors the one entity in the chain that can actually be authenticated. And decision history: every score, every flag, every enforcement action, and every appeal outcome, retained with attribution of which signals fired and whether a model or a human decided.

The recording party is the platform, unilaterally, at the point of ingestion and afterwards. That is the defining constraint of the approach. Distributors do not attest, generation tools do not watermark reliably, creators do not sign. The platform records what content and its handlers do, because it cannot rely on what anyone says.

5.2. How it Verifies

Verification is by convergence rather than by cryptography. A cryptographic system verifies by checking a signature. This layer verifies by checking whether independent evidence agrees: multiple detection methods, behavioural context, human review on the uncertain band, and appeals as an adversarial audit in which someone with private knowledge of a track's true origin can contest the reconstruction. Verification is probabilistic, continuous, and revisable.

That is simultaneously a weakness and a strength. Origin cannot be proved. Equally, the system cannot be defeated by a single forged credential, because an adversary has to fake consistent behaviour across signal classes rather than produce one artefact that verifies.

5.3. Failure Modes

Five failure modes are enumerated here to match the five in Section 4, because Section 6 pairs them.

Adversarial post-processing. Audio-level detection can be defeated by re-recording, filtering, and perturbation. Its structural weakness is that it examines the artefact in isolation, which is precisely the setting in which optimisation against a known decision boundary is most effective [14], and the generalisation results in [6], [7], and [9] show how narrow the margin is even without a deliberate adversary.

Patience. Behavioural signals can be defeated by slow uploads, varied metadata, and imitation of legitimate long-tail artists. A sufficiently careful adversary looks normal.

Stolen or fabricated identity. Identity anchoring assumes identity is expensive. Accounts can be bought, taken over, or aged deliberately before use, and the Sybil formulation makes clear that this is the expected behaviour of an open system rather than an aberration in one [17].

Definitional ground truth. There is no oracle for whether a given track was generated, partly because the question is partially definitional at the boundary. A behavioural provenance system therefore cannot report an oracle-verified accuracy figure, and any such system that claims one should be distrusted.

Probabilistic output as an adoption barrier. A confidence score is an invitation to argue. This is not a modelling failure; it is a failure mode of the layer as deployed infrastructure, and Section 6.4 treats it as the hardest one.

6. Composing the Two: Coverage Where Each Is Weak

The two layers are not competing approaches to the same problem. They are a layered system in which cryptography, where it exists, collapses the problem to a lookup, and behavioural inference covers the currently enormous remainder. Table 1 states the comparison on the properties that determine how they compose.

One distinction runs through the whole of this section and belongs at the front of it. The composition that was operated is the one internal to the behavioural layer, in Section 6.1: three signal classes chosen so that no two of them fail under the same conditions. The composition of the behavioural layer with a cryptographic one, in Section 6.2, was not operated and could not have been, because no cryptographic provenance layer was present in the supply chain this account draws on. It is offered as a proposal that follows from the same design rule, and it is analytical throughout. Figure 2 marks the two directions accordingly, Figure 3 states the routing logic of Section 6.3, and Section 6.4 reports the constraint that turned out to bind hardest in practice.

Table 1. Cryptographic and behavioural provenance compared, on the properties that determine how they compose.

Table 1. Comparative Analysis of Cryptographic Provenance and Behavioural Reconstruction

Property	Cryptographic provenance	Behavioural reconstruction
Verification mechanism	Check a signature; the answer is deterministic	Check whether independent evidence agrees; the answer is a confidence
What it establishes	Origin, for covered content	How content and its handlers behave, and whether anyone credible contests it
Cooperation required	The generating tool and every intermediary must participate	None; the platform observes at its own front door
Coverage today	Content produced from now on, by cooperating tools, through compliant pipes	The existing catalogue and every new arrival, on the author's account of the system he operated; no coverage figure was released
Cost profile	Cheap per item once deployed, expensive to deploy across an industry	Expensive per item and continuously, cheap to start unilaterally
Effect of absence	Absence carries no information while absence is the norm	A new arrival has no accumulated history yet, so absence is the starting state of every item and is resolved by accretion over time rather than read as a signal (Section 7.3)
Primary failure modes	Absent at source, stripped in transit, forged credential, key compromise, verification cost	Post-processing, patience, stolen identity, definitional ground truth, probabilistic output
Adversary's task	Obtain a recognised credential, or strip one	Fake consistent behaviour across orthogonal signal classes, indefinitely
Institutional acceptance	High; a signature verifies or it does not, though trust-list governance reintroduces some of the argument	Low by default; a score has to be argued for

6.1. The Design Rule and the Composition that was Operated

The composition rule has two clauses. Never let two signals share a failure mode. Never let one signal's confidence override another's alarm.

The first clause is what makes signals additive rather than merely numerous. Combining two signals that fail under the same conditions produces a system with one failure mode and more compute. Three signal classes were chosen against that test, and the chain they form is this paper's first contribution because it is the composition that was actually operated. Audio-level detection is gamed by adversarial post-processing, and its structural weakness is that it examines the artefact in isolation. It is covered by behavioural context, which does not care what the audio sounds like: upload cadence, catalogue structure, metadata patterns, distribution pathways. One track's audio can be laundered, but laundering the behaviour of an operation that exists in order to be industrial means abandoning the scale that made the fraud profitable, so the evasion cost lands on the adversary's

economics rather than on the signal. Behavioural context is in turn gamed by patience, since a sufficiently careful adversary looks normal, and is covered by identity anchoring: artist verification binds content to an authenticated entity with history and something to lose. An adversary can fake behaviour indefinitely, but accumulating legitimate identity is slow, costly, and burns on first enforcement, which makes it the closest thing a behavioural system has to a root of trust, not mathematically unforgeable but expensive to forge repeatedly. Identity itself is gamed by theft or fabrication, and is covered by decision history and the appeal loop, which make identity a running record audited by people with private knowledge rather than a one-time gate.

That three-part chain is the strongest claim in this paper, and it is a claim about a system that ran. Figure 2 sets out its pairings on the right-hand side and, separately and as a proposal, the cryptographic direction of Section 6.2 on the left.

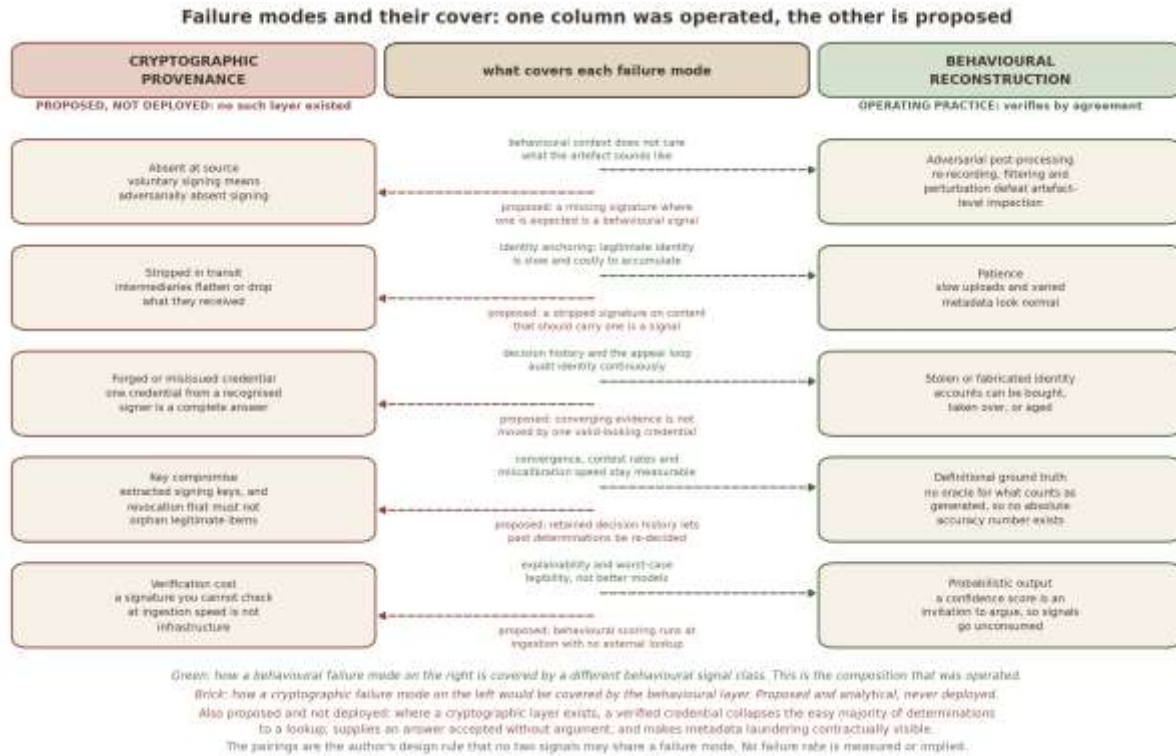


Figure 2. Comparative Failure-Mode Coverage: Cryptographic Provenance vs. Behavioural Reconstruction

Figure 2. Failure modes and their cover. The right-hand column and the green text are the behavioural composition that was operated. The left-hand column and the brick text are analytical and proposed: no cryptographic provenance layer was present in the supply chain this account draws on, so no cryptographic-to-behavioural coverage was deployed or observed. Pairings follow the author's design rule that no two signals may share a failure mode. No failure rate is measured or implied.

6.2. Composition with a Cryptographic Layer: A Proposal, Not a Deployment

Running the same test in the other direction produces the composition this paper proposes rather than the one it reports. It was not operated and could not have been, because no cryptographic provenance layer existed in the supply chain to compose with. What follows is reasoned from the failure modes in Section 4, from the standards literature, and from the architecture of [1]. It is not observed, and the paper does not want credit for it as though it were.

Cryptographic attestation, where present, would cover the behavioural layer's costliest weaknesses. It collapses the easy majority of determinations to a lookup, which is what would free inference capacity for the hard residue. It supplies an answer that is institutionally accepted without argument, which speaks to the adoption failure mode of Section 6.4 more directly than any modelling improvement does. And an attested chain of custody would make metadata laundering contractually visible rather than merely suspicious, which addresses the transit problem of Section 3 at its source instead of inferring it at the destination.

The corollary runs in the direction the author can speak to, and it is the one genuinely bidirectional element here. Each layer's failure becomes the other layer's input. A stripped signature is a cryptographic failure. A stripped signature on content from a supply chain that should have carried one is a behavioural signal, and a durable credential that cannot be recovered from a repository for content that should have one is the same signal a level down. The layers do not merely cover each other's gaps; the evidence of one layer's defeat is admissible evidence in the other. That property is why the behavioural architecture is the right one to have built on the assumption that cryptographic provenance eventually arrives, and it is why this subsection is a continuation of the design rule rather than a departure from it.

6.3. The Routing Rule

The second clause of the design rule determines what happens to a determination once the signals are in. Convergence earns automation. Divergence earns a human. An adversary who defeats any single layer does not exit the system; they promote themselves into a queue where a different layer is waiting.

Divergence is treated as informative in its own right rather than as noise to be resolved by weighting. When independent evidence and an external claim diverge sharply, either the inference is wrong, which is worth knowing, or the claim is fraudulent, which is worth knowing more. Both outcomes justify the cost of a person. Figure 3 shows the full path, including the property that makes the whole arrangement tolerable: every determination is retained with its evidence and can be re-decided.

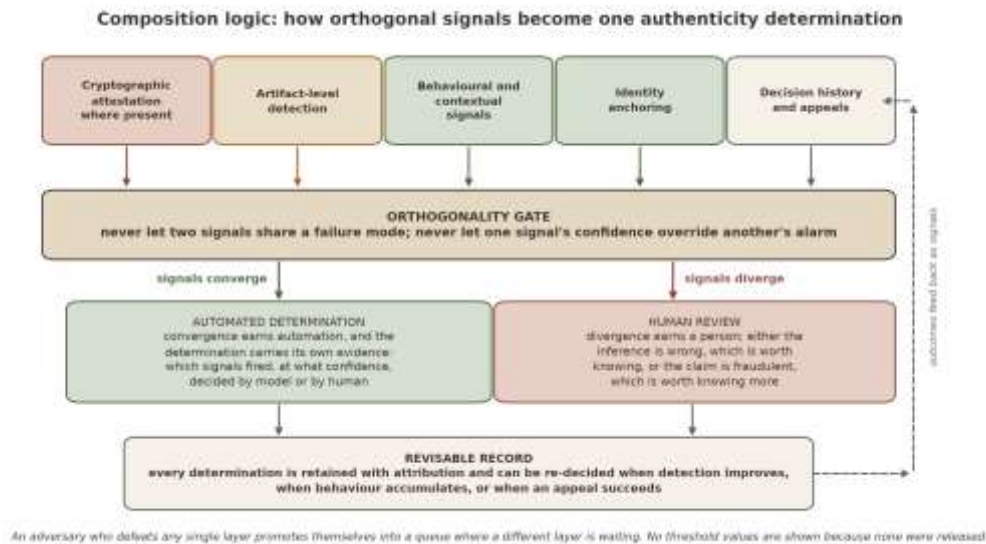


Figure 3. Composition Logic for Multi-Signal Authenticity Determination

Figure 3. The composition logic that produces an authenticity determination. No threshold values are shown because none were released.

6.4. The Composition Constraint Nobody Budgets for

The hardest obstacle in deploying this architecture was not technical. The models, the pipelines, and the scale were tractable engineering. The hard problem was getting an organisation to act on probabilistic provenance.

A cryptographic system, had one existed, produces answers people accept without argument. A behavioural system produces confidence scores, and a confidence score invites argument. Every downstream consumer of the signals, including royalties, editorial, recommendations, and enforcement, had a different tolerance for being wrong, a different cost of error, and a different threshold at which it would allow an inference to change what its system did. The early failure mode was signals that were computed and not consumed. Detection without enforcement is a dashboard, not a trust system.

Three things addressed it, and all three are architectural choices rather than modelling ones.

The strictest consumer set the bar. Rather than shipping a general-purpose score and leaving each team to interpret it, the consumer with the highest cost of error was identified and precision was engineered to that requirement. That inverted the adoption dynamic. Instead of every team auditing whether the signals were good enough for its purposes, the hardest customer's acceptance became the credential that let everyone else adopt without re-litigating. Meeting the strictest bar once is cheaper than negotiating a separate bar with every consumer indefinitely.

Every decision was made attributable. Adoption stalls when a consuming team cannot answer "why was this flagged?" to its own stakeholders. Every score carried its

evidence, including which signals fired, at what confidence, and whether a model or a human decided, so downstream teams were not asked to trust a number; they were handed a case. The audit trail turned out to be an adoption feature more than a compliance one. Teams will act on inference they can explain and will not act on inference they cannot.

Errors were priced in the consumer's own units. A false positive rate stated as a percentage was restated in the consuming team's unit of harm, as a count of artists wrongly flagged per month rather than a rate. Teams do not adopt a system because its precision is high. They adopt it when they can see its worst case and find it survivable.

The general lesson is that the binding constraint on behavioural provenance is institutional willingness to act on probability, not detection accuracy, and that willingness is not won with better models. It is won with explainability, worst-case legibility, and one hard-won reference consumer.

7. The Unsigned Majority

Content with no provenance is not an exception to be handled. It is the operating environment. Effectively everything arriving at a music platform today carries no verifiable history, so the question "what do you do about unprovenanced content" is really the question "what does the whole system do". The answer has four parts, shown in Figure 4.

7.1. Default to Inference, not to Trust or Distrust

Both naive policies fail, and they fail asymmetrically, which is why the choice between them is not a matter of risk appetite.

Trust by default hands the platform to industrial fraud, because the fraud is industrial precisely because the default is permissive. Distrust by default punishes the legitimate long tail, which has no way to prove itself, is unprovenanced for the same reasons everyone else is, and constitutes most of

the catalogue. Distrust by default also fails on its own terms, because the adversary with a forged credential is unaffected by a policy that penalises the absence of credentials.

So absence of provenance triggers measurement. Every item is scored on audio signals and behavioural context, and the score, not the missing paperwork, determines what happens next. The fairness principle underneath is precise and it is the load-bearing claim of this section: lacking provenance is not itself evidence of anything, because when effectively all content lacks it, its absence carries no information. What carries information is what the content and its uploader do. This is the same asymmetry identified in Section 4, applied as policy: a verified signal is evidence, a missing signal is a base rate.

7.2. Let Cheap Signals Settle Most of it

Most unprovenanced content is, on the author's account, unremarkable: a real person's demo, a small label's back catalogue, a hobbyist's upload. Behavioural and contextual signals settle the bulk of the catalogue confidently without expensive analysis, again on his account; no proportion was measured for publication and none should be read into the wording. The deep inspection budget, meaning expensive models and human review, goes only where cheap signals leave genuine uncertainty. Cluster-level review compounds the effect, since one decision can resolve an entire upload campaign rather than one item at a time.

This is an economic constraint stated as an architectural one. A universe of unprovenanced content can be processed only by refusing to treat every item as equally suspicious. Uniform suspicion is not merely unfair, it is unaffordable, and the two failures arrive together.

7.3. Let Provenance Accrete

Unprovenanced is not a permanent state, and this is the part of the policy that is least represented in the literature. Content that arrives with no history builds one: stream behaviour over time, catalogue consistency, verification events, survived appeals, enforcement history. A track that has been in the catalogue for two years behaving the way music behaves is not epistemically equal to one that arrived this morning, even though both arrived unprovenanced.

The platform effectively issues provenance retroactively, from observed behaviour. This is weaker than cryptographic history in every respect except one that matters operationally: it cannot be forged quickly. Accumulated behavioural history has the same economic property that makes identity anchoring work, which is that faking it costs time that the adversary's business model does not have. It compounds, and it compounds in favour of the honest participant, which inverts the trust burden identified in Section 1.

7.4. Keep Every Decision Revisable

Because inference substitutes for knowledge, no decision on unprovenanced content is final. New signals, improved detection methods, and appeals cause the system to re-decide. Retroactive scoring of the historical catalogue was this principle applied at catalogue scale: when the ability to infer improved, the past was re-inferred.

Revisability is what makes the whole posture defensible. A system that must decide without knowing, and that cannot revisit its decisions, converts every inference error into a permanent one. A system that can revisit them converts errors into a backlog, which is a materially different kind of problem.

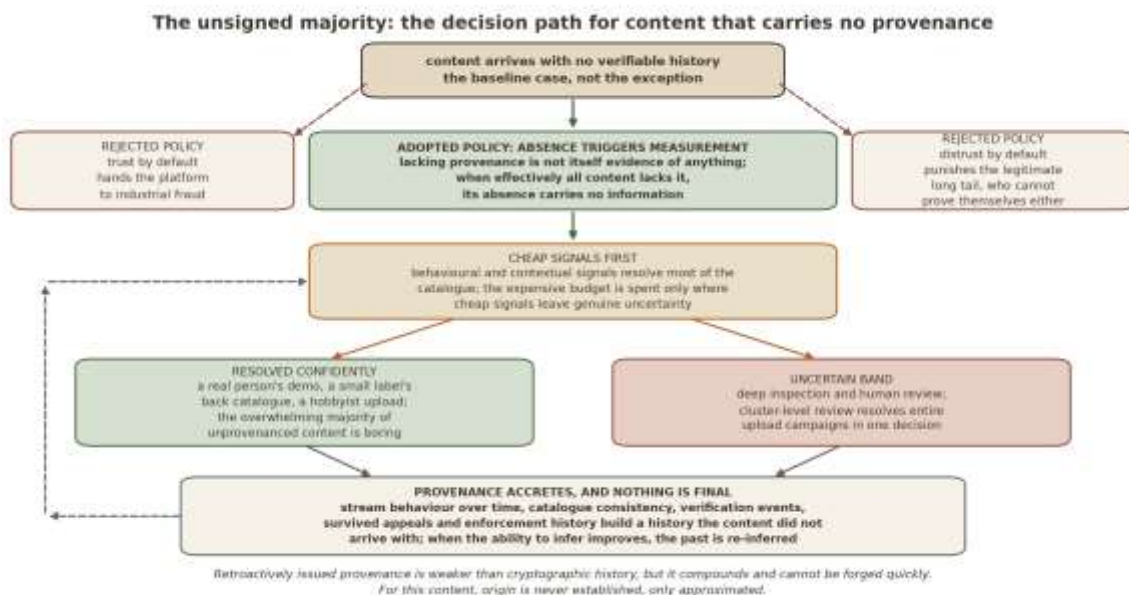


Figure 4. Decision Framework for Content without Verifiable Provenance

Figure 4. The unsigned-majority decision path. Structural policy only; no coverage percentages or review ratios are shown, as none were released.

7.5. The Honest Limit

None of this establishes origin. For unprovenanced content the platform never learns where the content came from. It learns, with accumulating confidence, how the content behaves and whether anyone credible contests it. That is the actual state of the art and papers in this space should say so plainly. For most content that exists, provenance is not recovered. It is approximated, continuously, at considerable cost.

8. Verifying Third-Party Provenance Claims

Almost every provenance claim in the system is one the platform did not generate: metadata asserting an artist name and ownership, distributor submissions asserting the right to distribute, artist assertions, and appeal statements asserting how a track was actually made. None of it arrives with cryptographic backing. The validation posture is to treat every external claim as testimony and to test it the way testimony is tested, against evidence, consistency, and the claimant's stake. Figure 5 shows the four tests and where each of them breaks.

8.1. Consistency against Observed Behaviour

A claim validates or degrades according to whether the world around it agrees. Metadata claiming an established artist is checked against that artist's verified profile and catalogue history. A claimed release pattern is checked against actual upload behaviour. Claims do not have to be provably true to be usable; they have to be consistent with everything else observable, and inconsistency is itself a strong signal. On the author's account, fraud characteristically fails not on the forged claim but on the halo of behaviour around it that does not match; no frequency was measured for publication. This is the provenance application of the coordinated-behaviour result in [18]: the individual assertion may be unfalsifiable while the pattern of assertions is not.

This test breaks when an adversary is patient enough to make the halo consistent as well, which is the same failure mode as Section 5.3's patience, arriving one level up.

8.2. Weight Claims by the Claimant's Stake and History

An assertion from a verified artist with years of catalogue and an authenticated identity is not epistemically equal to one from an account created last week. The reason is not that verification is unforgeable. It is that accumulated legitimate identity is expensive to build and burns on first

misuse. Claimants earn evidentiary weight over time and lose it in a single event. Track record is the collateral behind testimony, and the Sybil result [17] is the formal statement of why collateral is required at all in a system where identities are cheap to mint.

This test breaks when identity is stolen rather than built, or aged deliberately before use, which is why it cannot be the outermost test.

8.3. Adversarial Validation through Appeals

The strongest external claims received are appeals, in which someone asserts, with private knowledge, that a determination about their own track is wrong. Appeals are validated the same way as any other claim, by weighing them against detection evidence, behavioural history, and identity, with human review. The property that makes appeals uniquely valuable is that the outcome feeds back. A claim that survives adversarial review retroactively validates the claimant. A claim that collapses degrades every future claim from that source. Appeals are therefore not a customer service function bolted onto an enforcement system; they are the only channel through which the system receives information from parties who actually know the answer.

This test breaks when nobody with private knowledge ever contests, because then nothing is audited, and silence is indistinguishable between a correct determination and an unnoticed one.

8.4. Never Let a Single Claim Override Converging Evidence

External claims are inputs to the confidence calculation, never trump cards over it. When testimony and independent evidence diverge sharply, that divergence routes to a human, for the reason given in Section 6.3: either the inference is wrong or the claim is fraudulent, and both are worth the cost of finding out.

This test breaks when review capacity is the binding constraint on how much divergence can actually be read, which is a resourcing limit rather than a design limit, and it is real.

8.5. The Honest Limit

External claims cannot be validated to certainty. They can only be priced. A claim's value in the system is a function of its consistency with observation and the cost the claimant paid to be able to make it. That is testimony jurisprudence implemented as infrastructure, and until upstream attestation exists, it is the only validation available.

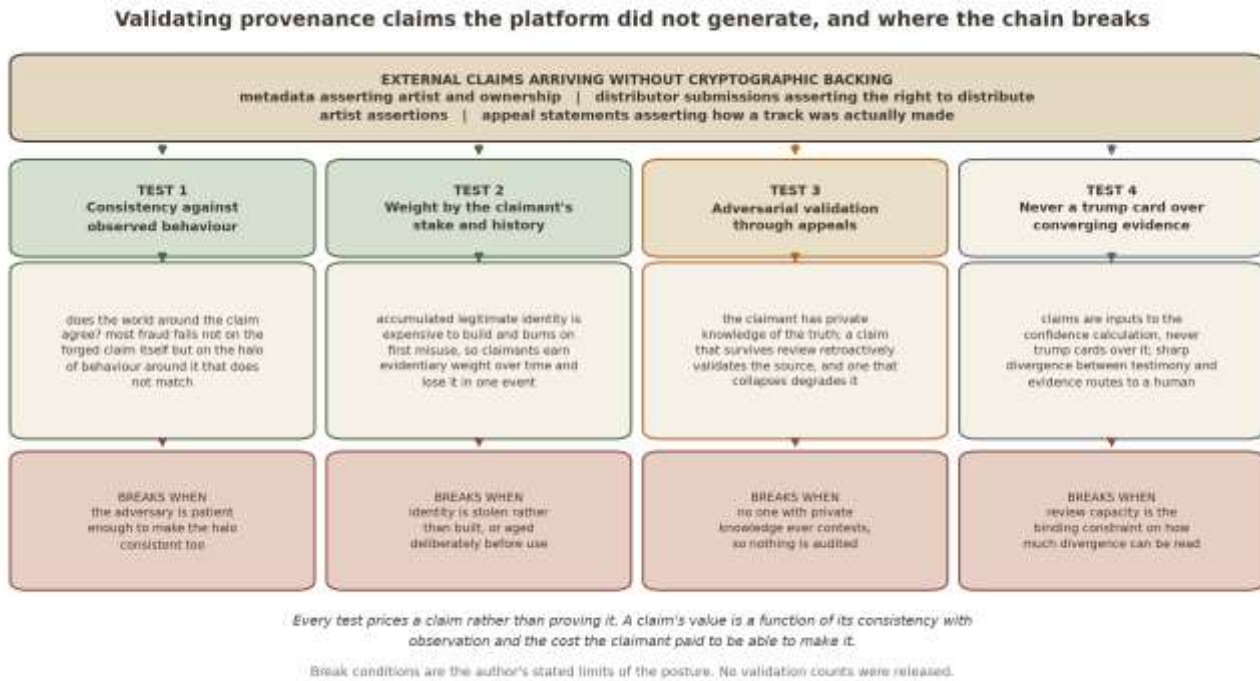


Figure 5. Validation Framework for External Provenance Claims and Chain-Break Analysis

Figure 5. The third-party claim trust chain and where each test breaks. Break conditions are the author's stated limits of the posture; no validation counts were released.

9. What Must Change across the Industry

Provenance at scale is not blocked on cryptography. The mathematics has existed for years, the standards exist, are versioned, and are deployable [3], and systems work has shown them carried into demanding delivery contexts [4]. What is missing is everything around the mathematics. Five

changes are named here in the author's order of how much they matter, and Table 2 states them against the lever that would actually move each one.

Table 2. What must change for provenance to work at scale, in the author's priority order.

Table 2. Strategic Change Requirements, Barriers, and Enabling Levers for Provenance Adoption

Rank	Change required	Why it is currently blocked	The lever that would move it
1	Generation tools sign at the source, by default, with no opt-out easier than compliance	The tool's customer is the uploader, not the ecosystem, so labelling reduces the value of the product to the person paying for it	Regulatory force, or platform-side economic pressure; voluntary schemes are adversarially absent by construction
2	The supply chain preserves what it receives	Preserving history costs something at every hop and dropping it costs nothing; standards define the container but assign no consequence	Liability and contract, not standards; platforms are the paying customer of distributors and can require attested chain of custody as a condition of ingestion
3	Absence of provenance starts to cost something, gradually	Unprovenanced content enjoys the same defaults as attested content, so attestation is pure cost with no return	Differential treatment, sequenced as reward presence first and penalise absence only once tooling is universal and free
4	Verification is cheap and revocation actually works	Verification cost at ingestion speed is treated as an implementation detail, and revocation is unglamorous operational work	Funding the boring parts: key rotation, revocation that does not orphan legitimately signed history, verification at catalogue throughput
5	The industry accepts that behavioural inference is permanent	The prevailing narrative is that provenance replaces detection, which is commercially convenient and false	Reframing the goal: provenance collapses the easy majority to a lookup, which concentrates inference capacity on the hardest residue

Two of these deserve expansion, because they are the ones where the analysis here departs from the standard account.

Liability, not standards, is the fix for the transit problem. Standards bodies can define the container format indefinitely. Until a distributor bears a consequence for delivering content whose claimed history it laundered, the pipes will keep washing metadata. The lever is contractual and economic rather than technical, and it already sits with the platforms, who are the paying customers of the distribution layer and can require attested chain of custody as a condition of ingestion. That leverage was available and was not used. It is the single highest-impact move available today and it requires no new technology.

The sequencing of differential treatment is not a detail. The endgame has to be that signed content receives faster decisions, better monetisation defaults, and lighter enforcement friction, because otherwise attestation never pays for itself. But the transition matters enormously, because most legitimate content, including the entire back catalogue and every independent artist without tooling, is unprovenanced through no fault of its own. Penalising absence too early builds a system that punishes the long tail in order to inconvenience fraudsters who will forge credentials anyway. The order is: reward presence first, penalise absence only after the tooling is universal and free. Getting this order wrong is the most likely way for a well-intentioned provenance mandate to do net harm.

The final point is the one most often left out. Even in the best case, cryptographic provenance covers content going forward, from cooperating tools, through compliant pipes. It never covers the existing catalogue, hostile tools, or stripped signatures. The target architecture is therefore not provenance replacing detection. It is provenance collapsing the easy majority to a lookup, which concentrates inference capacity on the shrinking, hardest residue. Platforms should want provenance adoption precisely because it makes their behavioural systems sharper rather than obsolete. Anyone selling a future in which the inference layer gets switched off is selling something.

The one-sentence version: the technology is ready, the incentives are absent, and the leverage to fix the incentives sits with the platforms, which are currently paying for industrial-scale inference instead of using their purchasing power to make the problem smaller.

10. Limitations

The limitations of this paper are substantial and are stated without hedging.

One platform, one content domain. The account is practitioner experience from a single large-scale music streaming platform. Music has properties that shaped every conclusion here, including a supply chain that concentrates through a relatively small number of distribution pathways, which is what makes unilateral instrumentation at the front door tractable, and an identity layer in artist verification that provides an anchor other domains may lack. The architecture may transfer; the claim that it transfers is not tested.

No controlled evaluation. There is no experiment in this paper. No baseline was held out, no ablation was run on the composition rule, and no comparison against an alternative architecture was performed. The design rules in Section 6 are operating principles that were applied, not hypotheses that were tested against a control.

No released operational metrics. The author declined to release raw operational data. Accordingly this paper reports no coverage percentages, no precision or recall figures, no false positive rates, no appeal overturn rates, and no human review ratios, and the figures and tables report structure rather than measurement. The scale statements that do appear, namely a catalogue of more than 250 million tracks with retroactive scoring of historical content and tens of thousands of new tracks scored daily at ingestion, together with dozens of distribution pathways covered and no cooperative provenance partners, are the author's account of coverage and are not accompanied by an exact count of distributors or aggregators, which the author declined to estimate.

No absolute accuracy claim is available, by construction. Whether authenticity determinations were correct in an absolute sense was never measured, because ground truth for generation is partially definitional and no oracle-verified accuracy number existed. What was measured operationally was convergence, contest rates, and how quickly miscalibration surfaced. This is a limitation of the approach and not only of this paper, and any behavioural provenance system reporting an absolute accuracy figure should be read sceptically for the same reason.

Deterrence is unmeasured and unmeasurable. Fraud that never arrived because the system existed is likely a real effect and cannot be quantified. Papers in this space should stop implying otherwise.

One of the two composed layers was absent from the deployment. This is stated in the abstract, in Section 1, in Section 1.2, and at the head of Section 6, and it is restated here because it bounds every claim in the paper. No cryptographic provenance layer existed in the supply chain this account draws on. Section 6.1 reports a composition that ran. Section 6.2 proposes one that did not and could not have. Section 4 is analytical throughout.

Boundary against the author's prior published architecture. The cryptographic layer discussed in Section 4 is not this paper's contribution. Its specification, layering, threat model, and research agenda belong to [1] and are cited rather than restated. This paper's contribution is confined to the behavioural layer, its internal composition, the proposed composition with a cryptographic layer, the policy for unprovenanced content, the validation of third-party claims, and the industry analysis in Section 9.

11. Conclusion

Content provenance is discussed as though the remaining work were cryptographic. It is not. The standards

exist, they are deployable, and the author's prior work set out the deterministic architecture in which authenticity is established at generation and verified independently downstream [1]. The unfinished work is everywhere else: in supply chains that strip what they receive because nothing obliges them to preserve it, in the absence of any cost to absent provenance, and above all in the overwhelming majority of content that carries no verifiable history and still has to be decided about today.

This paper argues that the answer to that majority is a second provenance layer with its own standing rather than a stopgap. Behavioural reconstruction records what content and its handlers do, verifies by convergence rather than by signature, and composes internally under two rules: no two signals share a failure mode, and no signal's confidence overrides another's alarm. That composition is the one this account can report, because it ran. Extending the same two rules across a cryptographic layer is a proposal here rather than a result, since none was available to compose with, and its most useful consequence is that each layer's failure would become admissible evidence in the other, a stripped signature on content that should have carried one being itself a behavioural signal.

For the unsigned majority the operating policy is that absence triggers measurement rather than trust or suspicion, because when effectively all content lacks provenance the absence of provenance carries no information. Cheap signals settle most of the catalogue, the expensive budget is reserved for genuine uncertainty, provenance accretes retroactively from observed behaviour, and no determination is final. For claims the platform did not generate, testimony is priced rather than believed, against consistency, claimant stake, and adversarial contest.

None of this recovers origin, and the paper does not claim it does. What it claims is narrower and more useful: that the target architecture is not provenance replacing inference, but provenance collapsing the easy majority to a lookup so that inference can be concentrated on the hardest residue, and that the levers holding this back are liability, sequencing, and unglamorous operational funding rather than mathematics.

References

- [1] B. Kandati, "From Detection to Provenance: A Deterministic Architecture for Authenticating AI Generated Content," *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, vol. 7, pp. 238-240, 2026, doi: 10.63282/3050-9262.ijaidsm-l-v7i1p139.
- [2] B. Kandati, "The Importance of AI-Generated Content Detection in the Future: Societal, Ethical, and Policy Implications," *Journal of Advances in Developmental Research*, vol. 17, no. 1, 2026, doi: 10.71097/ijaidr.v17.i1.1701.
- [3] Coalition for Content Provenance and Authenticity, *Content Credentials: C2PA Technical Specification*, version 2.4. [Online]. Available: https://spec.c2pa.org/specifications/specifications/2.4/specs/C2PA_Specification.html. Accessed: Aug. 10, 2026.
- [4] M. Mesa-Simón, A. Escobar-Molero, B. Sáez-Mingorance, D. P. Morales, J. A. Álvarez-Bermejo, and F. J. Romero, "Enabling Live Video Provenance and Authenticity: A C2PA-Based System with TPM-Based Security for Livestreaming Platforms," *IEEE Transactions on Multimedia*, pp. 1-12, 2026, doi: 10.1109/TMM.2026.3679049.
- [5] L. Verdoliva, "Media Forensics and DeepFakes: An Overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, pp. 910-932, 2020, doi: 10.1109/JSTSP.2020.3002101.
- [6] S.-Y. Wang, O. Wang, R. Zhang, A. Owens, and A. A. Efros, "CNN-Generated Images Are Surprisingly Easy to Spot... for Now," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 8692-8701, doi: 10.1109/CVPR42600.2020.00872.
- [7] R. Corvi, D. Cozzolino, G. Zingarini, G. Poggi, K. Nagano, and L. Verdoliva, "On The Detection of Synthetic Images Generated by Diffusion Models," in *ICASSP 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1-5, doi: 10.1109/ICASSP49357.2023.10095167.
- [8] X. Liu, X. Wang, M. Sahidullah, J. Patino, H. Delgado, T. Kinnunen, M. Todisco, J. Yamagishi, N. Evans, A. Nautsch, and K. A. Lee, "ASVspoof 2021: Towards Spoofed and Deepfake Speech Detection in the Wild," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 31, pp. 2507-2522, 2023, doi: 10.1109/TASLP.2023.3285283.
- [9] S. Oyucu, B. Arslan, and S. Sagioglu, "A System-Level Analysis of Cross-Domain Generalization Failure in Audio Deepfake Detection," *IEEE Access*, vol. 14, pp. 112734-112754, 2026, doi: 10.1109/ACCESS.2026.3716180.
- [10] Z. Yuan, L. Li, Z. Wang, J. Jiang, and X. Zhang, "Watermark Removal Attack Against Text-to-Image Generative Model Watermarking," *IEEE Signal Processing Letters*, vol. 32, pp. 1470-1474, 2025, doi: 10.1109/LSP.2025.3554514.
- [11] S. Kovačević, E. Nešović, K. Pavlović, P. Nedić, and I. Djurović, "DeepMark Benchmark: Redefining Audio Watermarking Robustness," *IEEE Access*, vol. 14, pp. 62031-62044, 2026, doi: 10.1109/ACCESS.2026.3685903.
- [12] P. Cano, E. Batlle, T. Kalker, and J. Haitsma, "A Review of Audio Fingerprinting," *Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology*, vol. 41, no. 3, pp. 271-284, 2005, doi: 10.1007/s11265-005-4151-3.
- [13] E. Drott, "Counterfeiting Attention in the Streaming Economy: Spam, Click Fraud, and Fake Artists," in *Streaming Music, Streaming Capital*, Duke University Press, 2023, pp. 144-192, doi: 10.1215/9781478027874-005.
- [14] N. Carlini and D. Wagner, "Towards Evaluating the Robustness of Neural Networks," in *2017 IEEE*

- Symposium on Security and Privacy (SP)*, 2017, pp. 39-57, doi: 10.1109/SP.2017.49.
- [15] Z. Newman, J. S. Meyers, and S. Torres-Arias, "Sigstore: Software Signing for Everybody," in *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, 2022, pp. 2353-2367, doi: 10.1145/3548606.3560596.
- [16] Y. Liu, W. Tome, L. Zhang, D. Choffnes, D. Levin, B. Maggs, A. Mislove, A. Schulman, and C. Wilson, "An End-to-End Measurement of Certificate Revocation in the Web's PKI," in *Proceedings of the 2015 Internet Measurement Conference*, 2015, pp. 183-196, doi: 10.1145/2815675.2815685.
- [17] J. R. Douceur, "The Sybil Attack," in *Peer-to-Peer Systems*, Lecture Notes in Computer Science, Springer, 2002, pp. 251-260, doi: 10.1007/3-540-45748-8_24.
- [18] Q. Cao, X. Yang, J. Yu, and C. Palow, "Uncovering Large Groups of Active Malicious Accounts in Online Social Networks," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, 2014, pp. 477-488, doi: 10.1145/2660267.2660269