Evidence Is Not a Verdict

CoreBond Research Series — CBRS-008
August 2026
Todd Kirton
DOI: 10.5281/zenodo.22242222

Why valid evidence can still be insufficient for the decision that follows

1. The Evidence Passed. Now What?

A platform produces attestation Evidence. The measurements are signed, the signature verifies, the freshness mechanism is acceptable, and the Verifier determines that the measured state satisfies its appraisal policy. It produces an Attestation Result that the Relying Party can validate and consume. (Birkholz et al., 2023; Lundblade et al., 2025)

Everything worked.

The next question is harder.

What, exactly, did that successful process establish—and what is the Relying Party now justified in doing because of it?

It might admit a workload to an internal service. It might release a production signing key. It might authorize access to confidential data. Each decision can consume a technically valid result, yet each decision may depend on a different proposition.

That distinction matters more than simply calling one action higher risk than another. A measured-state appraisal may establish that a platform satisfied defined reference values at a relevant time. A key-release decision may additionally depend on current execution state, key-custody conditions, workload identity, authorization, or some other proposition the original appraisal never attempted to establish.

The evidence did not fail. The question changed.

Validity asks whether evidence satisfies the technical rules applicable to it. Sufficiency asks whether the available evidence supports the bounded proposition the decision actually requires.

Paper 7 examined authority: who was entitled to assert, appraise, recognize, delegate, set policy, or decide. Paper 8 begins after that question has been answered. Assume the authority is legitimate and correctly scoped. The remaining question is whether the evidence and the inferences built from it justify what happens next.

When is evidence sufficient for a trust decision?

2. The Easy Thesis Did Not Survive

The first version of this paper had an attractive answer: higher-consequence decisions should require stronger evidence.

That sounds sensible. It is also too simple.

NIST SP 800-63-4 already uses digital identity risk management to assess potential adverse impacts from failures in identity proofing, authentication, and federation. Those assessments feed the initial selection of

Identity Assurance Level, Authentication Assurance Level, and Federation Assurance Level, after which organizations can tailor controls for the actual environment. (National Institute of Standards and Technology [NIST], 2025a)

NIST's broader risk guidance is equally clear. SP 800-30 treats likelihood, impact, and resulting risk as core inputs to risk assessment. Consequence-sensitive security decision-making is not a missing idea. (Joint Task Force Transformation Initiative, 2012)

RATS also rejects a universal evidence threshold. RFC 9334 separates the Verifier's appraisal of Evidence from the Relying Party's appraisal of the resulting Attestation Result. A result that passes the Verifier's policy does not automatically satisfy every downstream policy or use case. (Birkholz et al., 2023)

The research therefore killed the broad thesis.

Paper 8 cannot honestly claim that modern trust architectures generally ignore consequence when selecting assurance. They do not. It cannot claim that risk-based evidence requirements are new. They are not. And it cannot reduce the problem to a rule that more serious decisions always need more evidence.

What survived is narrower.

Policy can legitimately determine that evidence is sufficient. But in a composed trust system, that decision may occur several semantic steps away from the original evidence. Evidence is collected. A Verifier appraises it. Claims may be transformed or summarized. A Relying Party applies local policy. A consequence follows.

Each step can be locally valid while the final system becomes unable to explain which upstream evidence supported which downstream conclusion.

That is the problem Paper 8 keeps.

3. Validity Is Not Sufficiency

Evidence can be valid without being sufficient.

A certificate can validate correctly while saying nothing about the current runtime state of the endpoint presenting it. An attestation result can be authentic while being too old for a particular use. An identity assertion can be valid while failing to establish authorization for the requested resource. A posture signal can accurately report one property while leaving another required property unmeasured.

They require the decision to depend on a proposition the evidence does not fully establish.

RATS makes this distinction architectural. Evidence is appraised by a Verifier under an Appraisal Policy for Evidence. The resulting Attestation Result is then consumed by a Relying Party under an Appraisal Policy for Attestation Results. The two stages exist because evidence appraisal and application-specific use are not the same decision. (Birkholz et al., 2023)

Freshness adds another dimension. RFC 9334 provides mechanisms and architectural considerations for determining whether Evidence or an Attestation Result is recent enough for the relevant use. Paper 5 reached the same broader conclusion: freshness is not a universal timer. What is current enough depends on what changed, what is being protected, and what decision follows.

Coverage matters as much as freshness. Evidence can directly establish one proposition while a receiving system silently treats it as support for several. A measured-boot appraisal may support a claim about defined boot state. It does not automatically establish current operator control, every aspect of runtime integrity, authorization for a business action, or custody of a protected key.

This is why evidence strength is often too vague a phrase. Evidence has dimensions: validity, relevance, freshness, coverage, provenance, and sometimes corroboration. Those dimensions do not collapse cleanly into one universal score.

Sufficiency is therefore not an intrinsic property attached to a piece of evidence.

It is a relationship between what the evidence establishes and what the decision requires.

4. The Same Result Can Meet One Decision and Miss Another

Consider a RATS-style system.

An Attester produces Evidence describing measured platform state. A Verifier appraises that Evidence and produces an acceptable Attestation Result. The Relying Party validates the result and combines it with an approved workload identity.

Suppose the requested action is access to a bounded internal service. The relying policy requires two propositions: that the workload identity is approved and that the platform satisfied a defined measured-state policy. The available evidence covers both. The Relying Party accepts.

Now the workload requests release of a production signing key.

The existing Attestation Result remains valid. But suppose key release additionally requires a proposition about current key-custody state or a specific execution condition not represented in that result. The original evidence cannot become sufficient for that missing proposition merely because the Relying Party wants to act.

The difference is not simply that a signing key feels more important than an internal dashboard.

The second decision asks the evidence to establish more.

The Relying Party may obtain additional evidence, require a fresher appraisal, use another trusted source, or deny the request. Those are policy choices. Paper 8 does not prescribe which one is correct.

The architectural requirement is smaller: the system should be able to distinguish the proposition the original Attestation Result actually supported from the additional proposition introduced by the key-release policy.

A third variation tests the boundary. In an operational environment, security controls coexist with safety, reliability, and performance requirements. NIST SP 800-82 Rev. 3 makes that coexistence explicit. (Stouffer et al., 2023)

That does not mean OT systems should casually operate on stale evidence. It means additional verification delay cannot be assumed to be harmless in every environment. A universal rule that always escalates evidence collection as consequence rises fails once the act of collecting, waiting, or rejecting carries its own consequence.

> **Evidence is sufficient for something.**

The proposition after for is the part an architecture cannot afford to lose.

5. More Evidence Is Not the Answer

Once a system recognizes that one result does not cover every proposition, an obvious response appears: collect more.

That is not a general solution.

Additional evidence can improve a decision when it adds relevant coverage, improves freshness, reduces a material uncertainty, or provides information from a meaningfully different trust or collection path. But signal count alone establishes none of those things.

Several signals sharing the same collection mechanism, software component, authority, or trust path may provide less independent corroboration than their number suggests. That is a caution, not a mathematical independence rule. Paper 8 does not claim a universal metric for evidence independence.

Likewise, three identity attributes may be irrelevant to the platform property being decided. Continuous telemetry may be fresh but noisy. A larger evidence bundle may expose additional private information while adding little to the proposition that matters.

Mature assurance frameworks do not generally define assurance as a raw count of signals. NIST specifies assurance requirements and controls appropriate to selected levels. RATS uses appraisal policy to determine which claims and reference values matter. (NIST, 2025a; Birkholz et al., 2023)

The useful question is therefore not how much evidence the system has.

It is what additional proposition, coverage, or failure resistance the next piece of evidence provides.

6. Being Wrong Has Two Directions

Security discussions naturally emphasize false acceptance. A malicious device is admitted. A fraudulent identity is accepted. An unauthorized request succeeds.

Those failures matter.

They are not the only engineering errors that matter.

NIST SP 800-63A-4 provides a concrete, domain-specific example. In automated validation of physical identity evidence, performance requirements account for both document false acceptance rate and document false rejection rate, with testing under conditions substantially similar to the operational environment. (NIST, 2025b)

That standard does not define a universal error model for every trust decision, and Paper 8 should not pretend it does. It demonstrates something narrower: mature evidence systems can explicitly measure harmful error in both directions.

Operational technology provides a separate boundary test. NIST SP 800-82 Rev. 3 emphasizes that cybersecurity requirements in OT must coexist with safety, reliability, and performance. A denial, delay, or unavailable dependency can therefore have consequences that differ sharply from those in ordinary enterprise access control. (Stouffer et al., 2023)

Again, the point is not that false rejection is always worse. It is not.

The point is that the decision rule cannot assume only one direction of error carries cost.

A defensible sufficiency rule asks not merely what happens if bad evidence is accepted, but also what happens if a legitimate request is rejected despite relevant evidence or required evidence cannot be obtained in time.

The weights are domain-specific.

The existence of both directions is not.

7. Evidence Changes on the Way to the Decision

The strongest reason to care about sufficiency appears in composed systems, where the component making the final decision may never see the original Evidence.

RFC 9711 makes this explicit for Entity Attestation Tokens. A Verifier can process claims and measurements, compare them with reference values, and produce Attestation Results that forward, modify, summarize, or supplement information according to policy. (Lundblade et al., 2025)

That transformation is useful. A Relying Party should not necessarily need every raw measurement or every device-specific appraisal rule. Specialized Verifiers exist precisely so downstream components can consume meaningful results rather than reimplement the entire appraisal process.

But abstraction creates a semantic boundary.

Suppose raw measurements support proposition A. The Verifier appraises them and emits result B. The Relying Party then combines B with local information and policy to conclude C, which authorizes consequence D.

A → B → C → D can be perfectly legitimate.

The architectural problem appears when the system can no longer distinguish what A actually established from what the Verifier added in B, what the Relying Party inferred in C, and why D followed.

This is not an argument that every Relying Party must retain every raw measurement. It is not a requirement that reference values, policies, and Evidence be stored forever. Paper 8 does not prescribe a retention format, retention period, database schema, or storage architecture.

The minimum requirement is semantic.

A later reviewer should be able to distinguish the proposition supported by upstream evidence from conclusions introduced by appraisal and local policy.

Paper 7 identified the authority boundary in this handoff. Paper 8 identifies the evidentiary boundary.

Locally correct components do not remove the need to understand the composition.

8. Decision Provenance

The plain-language requirement is decision provenance.

A trust system should preserve enough information about a consequential decision to distinguish four things: what evidence was accepted, what proposition that evidence supported, what appraisal or local policy added, and what consequence followed.

This does not mean preserving every raw signal. It means preserving the semantic path.

For convenience, this paper refers to that path as evidence-to-decision accounting. The phrase is shorthand, not standards terminology and not a claim of a new technical discipline.

A reviewer should be able to ask:

- What bounded proposition was the system trying to establish?
- Which evidence satisfied its applicable technical validation rules?
- Why was that evidence relevant to the proposition?
- Was it fresh and contextually applicable enough for this use?
- What parts of the proposition did the evidence cover, and what remained outside its scope?

- Did additional sources add meaningfully different information or failure resistance?
- Which appraisal or relying-party policy introduced additional inference?
- What consequence followed from the final decision?

Those questions separate observation from inference.

That separation matters because composed trust systems can amplify meaning without falsifying anything. A Verifier can truthfully report that Evidence satisfied its appraisal policy. A Relying Party can truthfully report that its local policy authorized an action. Yet the final authorization may depend on propositions that were never directly present in the original Evidence.

Decision provenance makes those semantic additions visible.

It does not tell the policy what threshold to choose. It tells the architecture not to erase the path by which the threshold was satisfied.

That is the difference between this argument and the tautology that policy decides sufficiency.

9. What This Does Not Solve

Decision provenance does not make evidence complete.

It does not eliminate uncertainty. It does not produce a universal confidence score. It does not prove that multiple evidence sources are independent. It does not define one freshness threshold. It does not tell every system when to fail open or fail closed.

It also does not replace risk management. NIST already provides mature mechanisms for assessing impact and selecting assurance. RATS already provides appraisal-policy boundaries. OT engineering already recognizes that security controls interact with safety and availability.

Paper 8 asks a narrower question: when a system says the evidence was sufficient, can it distinguish what the evidence actually supported from what later appraisal and policy added?

Sometimes that examination will show that the evidence did not clear the decision's sufficiency rule.

Paper 8 stops there.

How the system represents, communicates, and acts on the resulting unresolved state belongs to Paper 11.

10. Questions Worth Carrying Forward

Paper 8 began with a suspicion that higher-consequence trust decisions simply needed stronger evidence.

The research did not let that survive.

Mature systems already connect impact to assurance, risk to controls, and use case to appraisal policy. They also recognize that controls themselves create costs and that rejection can be harmful as well as acceptance.

The narrower conclusion is more useful.

Evidence is not a verdict.

It can be valid without being sufficient. The same valid result can satisfy one decision while failing to cover a proposition required by another. Evidence can be transformed before reaching the component that acts on

it. More signals do not automatically add relevant coverage. And a composed system can reach a legitimate final decision while making the evidentiary path to that decision increasingly difficult to see.

That gives trust architecture a practical accounting problem.

What did the evidence establish?

What did appraisal add?

What did local policy infer?

Why did the resulting decision justify this consequence?

Those questions do not demand a new risk formula. They demand that semantic boundaries survive composition.

Once the decision is made, however, another problem begins.

A trust decision happens at a moment. Systems rarely return to a blank slate afterward. Results are cached. State is remembered. Prior decisions influence later ones. Something persists.

Paper 9 asks what should.

References

Birkholz, H., Thaler, D., Richardson, M., Smith, N., & Pan, W. (2023). Remote ATtestation procedureS (RATS) architecture (RFC 9334). Internet Engineering Task Force. https://doi.org/10.17487/RFC9334

Joint Task Force Transformation Initiative. (2012). Guide for conducting risk assessments (NIST Special Publication 800-30, Revision 1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-30r1

Lundblade, L., Mandyam, G., O'Donoghue, J., & Wallace, C. (2025). The Entity Attestation Token (EAT) (RFC 9711). Internet Engineering Task Force. https://doi.org/10.17487/RFC9711

National Institute of Standards and Technology. (2025a). Digital identity guidelines (NIST Special Publication 800-63-4). https://doi.org/10.6028/NIST.SP.800-63-4

National Institute of Standards and Technology. (2025b). Digital identity guidelines: Identity proofing and enrollment (NIST Special Publication 800-63A-4). https://doi.org/10.6028/NIST.SP.800-63A-4

Stouffer, K., Pease, M., Tang, C., Zimmerman, T., Pillitteri, V., Lightman, S., Hahn, A., Saravia, S., Sherule, A., & Thompson, M. (2023). Guide to operational technology (OT) security (NIST Special Publication 800-82, Revision 3). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-82r3