

Higher-Categorical Semantics of Signed Agent Coordination and Conservative Protocol Evolution

Anonymous

24 October 2026

Abstract

We formalize signed agent coordination over existing providers. Witness semantics separates evidence, authority, and effects. Conditional results establish origin authenticity, budget conservation, replay safety, result uniqueness, and conservative evolution. Counterexamples delimit truncation, encryption, and offline revocation.

1. Scope and contribution

Assume existing custody, domain email validation, key expiry/renewal, transactional state, and audit providers. We study their composition across agents, daemons, cloud authorities, and virtual machines. A skill is an instruction-bearing adapter, not an enforcement mechanism.

The specification connects authenticated bytes, scoped authority, committed effects, and versioned refinement. Categorical structure organizes evidence and trace equivalence. Security is conditional on the stated assumptions; no deployment or novel cryptographic primitive is claimed.

Application messages crossing principal boundaries require asymmetric signatures; network transport also uses TLS. Encryption is optional at profile selection and mandatory if either policy requires it. Authenticated local bootstrap IPC establishes signing contexts.

2. Messages, identity, and assumptions

Definition 1 (bound request). A request is

$$r = (v, p, a, b, o, x, g, \gamma, t_0, t_1, n, k, d),$$

where v, p identify the wire version and profile; a, b are sender and intended recipient; o, x name operation and resource; g, γ name grant and generation; $[t_0, t_1]$ is its acceptance interval; n is a random message identifier; k is a logical idempotency key; and d binds payload bytes or an ordered artifact manifest. Each detached artifact entry includes hash algorithm, digest, size, and position.

An injective, domain-separated encoding $B(r)$ binds every field. Profiles fix encoding, bounded parsing, signature suite, key binding, and rejection of duplicate names or unknown critical semantics [3,5]. Responses bind the initiating request digest, status, and output digests.

Assumptions A1-A6. A1: signatures are EUF-CMA secure [6] and artifact hashes collision resistant. A2: expected identities are correctly bound to public keys by approved issuers or pinned

roots. A3: custody isolates private keys and upstream tokens, and the broker signs only operations permitted to the authenticated workload. A4: authoritative admission and result transitions are durable and linearizable within their coordination aggregate [2]. A5: verification and execution enforce the specified parser, sandbox, resource, destination, and current-generation rules. A6: time-based decisions use a stated clock-error bound, maximum lifetime, and revocation-freshness policy. Compromise or failure outside these assumptions is outside the corresponding theorem.

Email evidence must bind principal, action digest, destination, nonce, scope, and expiry. It does not prove workload-key possession, which is checked separately. Renewal requires current lifecycle authority.

Proposition 2 (conditional origin authenticity). For honest uncorrupted keys, an accepted request whose bound content was never signed by the authorized broker yields either a signature forgery, a hash collision, or a violation of A2-A5.

Proof. Exclude violations of A2-A5. Acceptance then uses the correctly bound key and exact validated encoding. If $B(r)$ was never signed, its valid signature is an EUF-CMA forgery. Otherwise the encoded fields equal a previously signed request. Different detached content with the same bound digest gives a collision. These exhaust the ways to change bound content. A union bound applies over keys and hash events. This proposition concerns origin of bytes, not freshness, permission, or truth of the payload. $\square$

3. Attenuating grants and resource accounting

Definition 3 (constraint order). A constraint is $c = (S, I, b, e)$: allowed action-resource-recipient tuples S , validity interval I , resource vector $b \in \mathbb{N}^q$, and minimum encryption level $e \in \{0, 1\}$. Define

$$c' \preceq c \iff S' \subseteq S, I' \subseteq I, b' \leq b, e' \geq e.$$

Vector inequalities are componentwise. This order expresses permission narrowing, including stronger confidentiality requirements.

A grant adds issuer, subject, audience, root, generation, and delegation depth. Chains require matching subject/issuer endpoints, valid signatures, decreasing depth, and narrowing constraints. Trust is directional and does not imply reverse or third-party grants.

Lemma 4 (attenuation). The terminal constraint of a finite valid delegation chain is below every ancestor constraint.

Proof. Inclusion, interval inclusion, componentwise inequality, and the reversed encryption order are transitive. Apply transitivity along the chain. $\square$

Two one-unit child grants can overspend a one-unit root. Therefore admissions charge a shared root balance or consume disjoint allocations already charged to it.

Theorem 5 (budget conservation). Let a root start with balance $B \in \mathbb{N}^q$. Each admitted job of cost c_i atomically requires $c_i \leq R$ and changes R to $R - c_i$. If duplicate admissions do not debit again, then

$$R_m + \sum_{i=1}^m c_i = B, \quad R_m \geq 0.$$

Proof. The assertion holds initially. A rejection or duplicate leaves both sides unchanged. A new admission subtracts and adds the same vector and preserves nonnegativity by its guard. Induction proves the result, including interleaved descendants under A4. This accounting model does not refund cancellation; a refund extension needs a separate conservation law. $\square$

Renewal cannot exceed current owner authority, revive revoked generations, or silently widen scope.

4. Witness semantics in an infinity topos

Fix an infinity topos $\mathcal{X}$, for example $\text{Fun}(C^{op}, \mathcal{S})$ for small evidence-compatible interface contexts C [1]. Reindexings preserve designated snapshots and trust bindings; expiry or revocation is not automatically such a reindexing.

Definition 6 (evidence and admission). Let $M \in \mathcal{X}$ be the object of bound requests with declared snapshot. For finitely many checks choose evidence morphisms $E_i \rightarrow M$: signature/key binding, recipient, freshness, grant, content, and execution policy. Define

$$W = E_1 \times_M \cdots \times_M E_s, \quad A = \text{im}(W \rightarrow M) \hookrightarrow M.$$

Here image means the effective-epimorphism/monomorphism factorization, equivalently relative (-1) -truncation in $\mathcal{X}_M$. An operational admission supplies actual checked data $w : T \rightarrow W$ over its request; a mere assertion of inhabitation is not an executable certificate.

Theorem 7 (conjunction and contextual stability). Writing $A_i = \text{im}(E_i \rightarrow M)$, there are canonical equivalences

$$A \simeq A_1 \cap \cdots \cap A_s, \quad u^*A \simeq \text{im}(u^*W \rightarrow T)$$

for every $u : T \rightarrow M$.

Proof. Each projection $W \rightarrow E_i$ makes the image of W factor through A_i , giving $A \subseteq \bigcap_i A_i$. Over the intersection, each pulled-back E_i maps effectively epimorphically onto the base. Their finite fiber product does so as well: pullback stability and composition of effective epimorphisms prove this by induction. Thus the reverse inclusion holds. Pullback preserves the effective-epimorphism/monomorphism factorization, giving the second equivalence [1, Section 6.2.3]. $\square$

This concerns relative images, not preservation of arbitrary pullbacks by truncation. Global points need not detect equivalences. Local inhabitation does not supply a chosen global execution witness.

Higher evidence can retain certificates and coherent comparisons. In spaces, the maps $* \rightarrow *$ and $B(\mathbb{Z}/2) \rightarrow *$ have equal images but inequivalent witness spaces, where $B(H)$ is the classifying space of a group H . Admission truth does not reconstruct witnesses; homotopies between cryptographic evidence must be explicitly supplied.

5. The transactional reference monitor

Definition 8 (authoritative state). For one coordination aggregate let

$$\sigma = (G, N, J, F, R, O, L).$$

Here G records grants/revocations, N consumed message identifiers and horizons, J immutable job digests/states/fences, F accepted results, R root balances, O an outbox, and L event-indexed causal audit evidence. Physical serialization order may also be retained.

Transitions are atomic and append audit/outbox evidence in the same transaction domain.

Admit. Verify a current witness. An existing key returns its receipt only for the same immutable input digest. A new key requires an unused nonce, active generation, freshness, and sufficient balance; consume the nonce, debit once, and create the job.

Lease. An eligible uncompleted job receives an increasing fencing number. Expiry and renewal are explicit; an earlier lease does not establish current authority.

Commit. Require the exact current attempt and unexpired lease, active grant generation, permitted result digest/schema/recipient, a noncancelled job, and no existing entry $F(k)$. Insert $F(k)$ and mark the job accepted atomically.

Cancel/revoke. Invalidate the job or grant generation and advance fencing. Accepted results remain historical. Explicit new authority cannot make old attempts current again.

Mutable authority is revalidated at the linearization point; a separate unchecked check/write sequence is nonconforming.

Theorem 9 (replay, uniqueness, and fencing). Under A4-A6: (i) replay creates no additional logical admission or debit; (ii) a logical key accepts at most one result; (iii) a stale attempt cannot commit; and (iv) after cancellation or revocation linearizes, no later commit can succeed under the cancelled job or revoked generation.

Proof. Linearize concurrent transitions. For (i), after the first admission the logical key already exists, so later admissions return its receipt without mutation or reject differing content. Reusing a consumed message identifier for a new key fails while that message could remain acceptable. Retain nonce records through its entire acceptance horizon, including clock error; thereafter the signed time guard rejects it. For (ii), the first insertion makes the absence guard false permanently. For (iii), reassignment makes the old fencing number unequal to the current number. For (iv), cancellation or revocation falsifies a required current-state guard. Atomicity prevents a check/write race in each case. $\square$

This gives one accepted effect, not exactly-once CPU execution. External effects need their own idempotency contract.

Proposition 10 (conditional progress). If authorization and capacity persist, a job is not cancelled, delivery and scheduling are fair, retries occur, and some eligible attempt eventually completes and reaches the authority within a sufficiently stable lease, then the job eventually has an accepted result or already has one.

Proof. Fair delivery reaches an eligible attempt. Its successful completion eventually reaches Commit while the guards remain true, unless an earlier commit has already populated $F(k)$. In the former case it commits; in the latter the desired result already exists. $\square$

6. Directed traces and certified commutation

Let T be the free category on legal monitor transitions, with states as objects and finite legal paths as arrows. Empty paths are identities and concatenation is composition. Resource balances,

grants, time, and fencing are in the states, so composition does not silently restore spent or expired authority.

Call two transitions independent only when their read/write dependencies, grants, root balances, nonce domains, jobs, and approval consumptions permit both orders, with the same resulting semantic state. Ordered physical log position is excluded from this coarse observation; its evidence remains retained.

Lemma 11 (commutation). Deterministic transitions with disjoint read/write dependencies commute, provided neither changes a guard read by the other.

Proof. Each transition reads values unchanged by the other and updates disjoint components. Both guards have the same truth value in either order, and both componentwise updates agree. Event-indexed audit records combine by union with the same causal dependencies. $\square$

Adjoin invertible 2-cells for certified squares and close under the strict 2-category laws, giving a small $(2, 1)$ -category. Nerves of its hom-groupoids give Kan mapping spaces; the homotopy-coherent nerve is an infinity category. Shared budgets or approvals normally prevent independence.

Proposition 12 (forgetting directed labels loses safety). Unlabelled groupoid completion cannot determine whether a protocol admits an unsigned commit.

Proof. The category with one object and its identity has contractible classifying space and no commit. The interval category $[1]$, with its nonidentity arrow labelled unsigned commit, also has contractible classifying space but contains such a commit. Their groupoid completions are equivalent after forgetting labels, despite their different safety properties. $\square$

7. Conservative evolution and provider substitution

Definition 13 (release and refinement). A release binds schema, policy, wire-profile, provider-contract, implementation, and conformance digests, together with authenticated parent manifest digests. These versions are distinct. The first trusted root is independently established; a digest alone authenticates no publisher.

For an existing action vocabulary, a conservative migration assigns new states to old states and each new transition to a finite legal old path, possibly empty. Sources, targets, initial states, immutable operation bindings, resource accounting, encryption obligations, and required evidence are preserved. This extends uniquely to a functor $Q : \mathsf{T}_{\text{new}} \rightarrow \mathsf{T}_{\text{old}}$. Certified 2-cells additionally map to certified old comparisons.

Theorem 14 (safety preservation). Let I be an inductive old-state invariant. If migration maps each new initial state into I and each new transition to a legal old path, then every reachable new state satisfies I after applying Q .

Proof. Apply induction to a new execution. The base holds by hypothesis. A mapped new transition is a finite old execution; induction on its length preserves I because I is inductive for old transitions. The functorial extension handles arbitrary concatenations. $\square$

New actions need explicit specification and authority. Incompatible profiles fail negotiation. Provider replacement requires the same simulation obligation, including failure, renewal, expiry, and revocation.

Parent links record lineage; authenticated update metadata additionally resists rollback and staleness, as addressed by TUF [9].

8. Confidentiality and revocation boundaries

An encrypted profile protects a signed object for designated recipients, using an approved authenticated-encryption binding such as JWE [4]. Verify the inner signature after decryption. Both profile selection and encryption requirements are signed. Signature keys and encryption keys have separate custody and lifecycle roles. Encryption does not hide plaintext from the executor that must use it, prove result correctness, or automatically provide forward secrecy.

A signed bearer token remains usable by its holder unless the resource server enforces a bound proof-of-possession key. OAuth’s current security guidance and DPoP address this distinction [7,8]. DPoP does not by itself authenticate arbitrary request bodies, so it does not replace the bound-request signature.

Theorem 15 (offline revocation limit). Suppose an isolated node receives no information about an authority after a partition. If the node accepts an operation in some no-revocation history, it cannot guarantee rejection of that operation in every otherwise identical history where remote revocation has already committed.

Proof. Choose two histories identical to the node through its decision, with revocation occurring only at the unreachable authority in one history. Its observations are identical. A deterministic node makes the same decision; a randomized node has the same acceptance probability under coupled randomness. Acceptance in the first therefore permits acceptance in the second. $\square$

Thus bound offline grants or revocation staleness, or fail closed. Immediate remote revocation and unconditional isolated availability are incompatible.

9. Runtime binding and reproducibility

The cloud owns authority; a VM runs bounded jobs. Local brokers reuse pinned providers and prevent workload impersonation or unrestricted signing. Same-VM daemons cannot establish secrecy from a compromised administrator [12].

On Cloudflare, transactional Durable Object storage can own an aggregate [10]. A committed outbox bridges to Queues, whose delivery is at least once [11]. No atomic transaction across arbitrary databases, object storage, email, and queues is assumed. Cross-aggregate invariants require an additional coordination protocol.

The accompanying standard-library checker explores two jobs, input variants, finite nonce/fencing domains, shared budget, cancellation, and revocation. It checks invariants and counterexamples for omitted root accounting, fencing, and atomic consumption. Verification outcomes are abstract; cryptographic security, deployed providers, and unbounded execution are not tested. The root-budget-one and root-budget-two explorations visit 162 and 962 states, respectively; all checked invariants hold. Bounds accompany the artifact.

10. Authorship, limitations, and references

Authorship and automation. The responsible human authorized anonymous submission. An AI assistant assisted with formalization, writing, references, typesetting, and bounded checking. No prior peer review is asserted.

The checker is bounded evidence, not a mechanization of higher-topos theory or verification of A1-A6.

- [1] J. Lurie, *Higher Topos Theory*, Annals of Mathematics Studies 170, 2009; author-maintained text, Sections 5.1 and 6.2.3. <https://www.math.ias.edu/~lurie/papers/HTT.pdf>
- [2] M. Herlihy and J. M. Wing, “Linearizability: A Correctness Condition for Concurrent Objects,” *ACM TOPLAS* 12(3), 463-492, 1990. <https://cs.brown.edu/people/mph/HerlihyW90/p463-herlihy.pdf>
- [3] IETF, RFC 7515, *JSON Web Signature*, 2015. <https://www.rfc-editor.org/rfc/rfc7515.html>
- [4] IETF, RFC 7516, *JSON Web Encryption*, 2015. <https://www.rfc-editor.org/rfc/rfc7516.html>
- [5] IETF, RFC 8725, *JSON Web Token Best Current Practices*, 2020. <https://www.rfc-editor.org/rfc/rfc8725.html>
- [6] S. Goldwasser, S. Micali, and R. L. Rivest, “A Digital Signature Scheme Secure Against Adaptive Chosen-Message Attacks,” *SIAM J. Comput.* 17(2), 281-308, 1988. <https://people.csail.mit.edu/rivest/pubs/GMR88.pdf>
- [7] IETF, RFC 9700, *Best Current Practice for OAuth 2.0 Security*, 2025. <https://www.rfc-editor.org/rfc/rfc9700.html>
- [8] IETF, RFC 9449, *OAuth 2.0 Demonstrating Proof of Possession*, 2023, Section 11.7. <https://www.rfc-editor.org/rfc/rfc9449.html>
- [9] The Update Framework, *Specification*, version 1.0.37, 2026. <https://theupdateframework.github.io/specification/latest/>
- [10] Cloudflare, *SQLite-backed Durable Object Storage*, accessed 2026-10-24. <https://developers.cloudflare.com/durable-objects/api/sqlite-storage-api/>
- [11] Cloudflare, *Queues Delivery Guarantees*, accessed 2026-10-24. <https://developers.cloudflare.com/queues/reference/delivery-guarantees/>
- [12] OWASP, *Secrets Management Cheat Sheet*, accessed 2026-10-24. https://cheatsheetseries.owasp.org/cheatsheets/Secrets_Management_Cheat_Sheet.html