Nick Clark · Published July 23, 2026 · PDF

The Anti-Evasion Charge on Governance-Address Removal

An autonomous agent that has repeatedly needed a human principal to lift a withholding can be required to obtain a second signature before its next restoration, drawn from a governance address the principal itself named. The obvious way to escape that requirement is to publish a new policy that quietly drops the named address. The anti-evasion charge on governance-address removal, disclosed in U.S. Provisional Application No. 64/117,812, closes that route. Where an agent admits a successor principal policy object that removes a governance address while the co-signature condition still subsists and within a declared window after the recurrence bound was met, the agent scores the admission as a deflection event, applies a declared modification to its scoped integrity vector, and records the removed address, the subsisting condition, and the window. The condition is not lifted. The removal is priced rather than given effect.

1. The Gap

A governance control is only as strong as its weakest amendment path. Consider an agent whose principal keeps lifting a withholding that does not stay lifted. The architecture responds by requiring the next restoration to carry a second signature, verifiable against a governance address the principal named in its own policy. That

requirement exists because prior restorations recurred: the same origin-equivalence class kept forcing the authorization gate back into the withheld state, and single-signature restoration was not resolving the underlying condition. The second signer is a check on restoration authority.

Now ask where the name of that second signer lives. It lives in the principal's own policy object, which the principal is free to succeed. In any system where the latest admitted policy is simply authoritative, the check evaporates the moment its own address is edited out. The party being checked holds the pen that names the checker. Publish a successor policy that omits the governance address, admit it, and under naive policy semantics the co-signature requirement no longer has an address to verify against. The obligation dissolves without a single hostile determination ever being confronted.

The exposure is sharpest at the moment the recurrence bound is satisfied, when the second-signature requirement has just attached and the incentive to shed it runs highest. A defense that can be deleted by the entity it constrains is not a defense. The gap is not that policy is mutable; policy should be mutable. It is that one well-timed mutation would let an agent keep the benefit of dropping its overseer while paying nothing for it.

2. Mechanism

The anti-evasion charge is built on top of the co-signature condition, so the condition has to be stated first. The persistent semantic agent (100) computes a recurrence count from its append-only lineage field (104): it pairs each return of the authorization gate (300) to the granting state by a principal-resolution object with the first subsequent write to the withheld state (310) that corresponds in both origin-equivalence class (200) and action class, and counts the corresponding pairs. Responsive to that count satisfying a declared bound, the agent conditions the next return of the gate to the granting state upon both a principal-resolution object and a co-signature over that

object, verifiable against a governance address named in the principal's own admitted policy object under the same signature-verification procedure used for the signed policy object (112). While the condition subsists, a return presented with the resolution object alone does not write the gate. Where the principal's policy names no address, the condition is inoperative from the start.

Against that backdrop, the anti-evasion trigger is a conjunction of three predicates, all of which the agent already holds the material to test. First, the admission occurs within a declared window following satisfaction of the recurrence bound. Second, the co-signature condition still subsists at the time of the admission. Third, the agent admits a successor principal policy object that removes a governance address named in the immediately preceding principal policy object. Only when all three hold does the mechanism fire. A policy that never named an address is not caught, because there is nothing to remove. A successor that adds an address, or that changes something unrelated, is not caught. A removal after the condition has already been cleared, or outside the declared window, is not caught. The charge is narrow by construction: it reaches the one edit that would otherwise defeat a live obligation.

When the conjunction holds, the agent scores the admission as a deflection event and applies a declared modification to its scoped integrity vector (106). It records the removed address, the subsisting condition, and the window into its append-only lineage field. "Deflection event" is the same accounting unit the architecture uses when an agent contracts its declared commitments while an assertion against it is pending; here the thing being contracted is not a value-scope tuple but the set of governance addresses, and the timing is what makes it a deflection rather than an ordinary policy revision.

Everything turns on what happens next: the condition itself is not lifted by the removal. This is the difference between forbidding an action and pricing it. The architecture does not reject the successor policy; the agent may author whatever policy it likes, and the successor is admitted as its current policy. What the removal does not get is operative

effect against the obligation it was aimed at. The co-signature requirement continues to bar the next restoration exactly as before, because it now traces to the immediately preceding policy that the lineage still records, and the subsisting condition is unaffected by the fact that its address was later edited out. The agent has spent a scoped-integrity-vector modification and a recorded deflection, and it still owes the second signature.

That symmetry is why this sits in the assertion-cost-symmetry family. Honoring the obligation costs a second signature from a named authority. Shedding it by policy edit costs a deflection event and an integrity modification and still leaves the obligation standing. The removal route is made strictly worse than obtaining the co-signature or letting the condition clear on its own terms.

Two structural properties matter. The test is entirely self-contained: everything the three predicates require, the immediately preceding principal policy object, the record that the recurrence bound was satisfied, the declared window, and the fact that the condition subsists, already lives in the agent's own append-only lineage field. No registry, directory, third party, or coordinating execution node participates in detecting the evasion. And the charge is recorded, not merely computed: the removed address, the subsisting condition, and the window are written into that same lineage field, so the deflection event and the integrity modification are entered on the record rather than left as a transient calculation.

3. Operating Parameters

The filed paragraph and its supporting cluster describe each of these as declared, and assign no numeric value to any of them. Deployment sets the values.

- **Recurrence bound.** The number of corresponding return-then-rewrite pairs that arms the co-signature condition. Until the recurrence count satisfies this declared bound, no second signature is required and the anti-evasion charge has no condition to protect.

- **The declared window.** The interval following satisfaction of the recurrence bound within which a governance-address removal is caught and priced. A removal outside this window is an ordinary policy revision and is not scored.
- **Subsistence of the co-signature condition.** A gating predicate rather than a tunable magnitude: the charge fires only while the condition is live. Clearing the condition legitimately, discussed below, removes the exposure.
- **The declared modification to the scoped integrity vector.** The magnitude applied to the scoped integrity vector (106) when the conjunction holds. It is declared, not derived from any asserting party's characterization of the removal.
- **Supporting intervals.** The recurrence count is itself computed over a declared retention interval and a declared correspondence interval, and a declared durability bound governs the legitimate clearing route. The filed cluster states each as declared without fixing its value.

4. Composition

Within the principal-metering cluster, this mechanism caps a single, specific escape. The cluster opens with the recurrence statistic, which detects that restorations are not holding, and the co-signature condition, which requires a second signer. It then provides a legitimate exit: a durability count over restorations not followed by a corresponding rewrite, which clears the condition once the principal has demonstrated durable restoration, even if the recurrence count has not yet expired. The recursive-metering embodiment extends the same logic upward, metering a co-signing authority whose signatures also fail to produce durable outcomes by the same statistic used against the principal.

The anti-evasion charge does not touch the legitimate exits. An agent whose principal earns its way out through the durability count, or waits for the contributing pairs to expire from the retention interval, pays nothing. What the charge prices is the exit that

is not earned: deleting the checker. The cluster offers two honest doors, demonstrated durability or expiry, and closes the dishonest one by making it a recorded, priced deflection that leaves the room locked anyway.

The mechanism also inherits vocabulary and quantities from elsewhere in the filing rather than inventing its own. Scoring a policy contraction under a pending obligation as a deflection event is the same move the architecture makes against an agent that narrows its declared value set while a conduct evaluation artifact is undetermined; the anti-evasion charge applies that established accounting to a different contraction, the removal of a governance address. The quantity it moves, the scoped integrity vector (106), is the same quantity that permitted-deviation consistency evaluation and reparation act on, so the charge lands where the agent's broader authorization posture is already tracked rather than in a bespoke counter. And the self-contained, registry-free character of the detection matches the filing's general stance: identity and history are derived from the agent's own append-only lineage, not from any external coordinator, so the evasion is caught with no party other than the agent involved.

5. Prior-Art Distinction

Several established families require more than one authority to approve a sensitive action. Each is a real technique, and none of them is asserted to practice what this mechanism does; the point is the structural feature each family lacks.

Multi-signature and threshold authorization, including M-of-N wallets, hardware-security-module co-signing, and threshold signature schemes, enforce that a transaction carries a required number of signatures. The set of eligible signers, however, is itself a configuration item: whoever controls the configuration can remove a signer, and the requirement drops to reflect the smaller set. There is no self-recorded cost for removing a signer at the moment the second signature is about to bite, and no notion that the prior requirement persists after the signer is gone.

Policy engines, role-based and attribute-based access control, and policy-as-code systems treat the latest admitted policy as authoritative. Narrowing a policy to remove an approval requirement takes effect immediately and silently; the removal is the intended semantics, not an event to be priced. Here the succession is admitted, but a removal within the declared window while the condition subsists is scored and denied operative effect, so the obligation outlives the edit.

Programmable-governance constructs such as timelocks and guardian vetoes impose a delay before a configuration change takes effect, or let a designated party block it. This mechanism imposes neither. The change is admitted at once; what follows is a recorded charge and a persisting obligation, computed from the agent's own append-only history with no on-chain registry or coordinating node in the loop.

Key rotation and certificate revocation offer the closest surface analogy, since a governance address behaves like a named verifying key. Standard public-key infrastructure lets a key holder rotate or revoke keys freely, and that freedom is normally unremarkable. The mechanism does not restrict rotation in general. It reaches only the narrow case where a named verifying address is removed within a declared window while a dependent co-signature condition is still live, and it prices the removal rather than preventing it.

6. Disclosure Scope

The mechanism described here is disclosed in U.S. Provisional Application No. 64/117,812, Section 10.3, at paragraph [0374], with its supporting predicate cluster at paragraphs [0370] through [0373]. What is disclosed is the anti-evasion charge on governance-address removal: the three-part trigger requiring a removal of a named governance address, within a declared window following satisfaction of the recurrence bound, while the co-signature condition subsists; the scoring of that admission as a deflection event; the application of a declared modification to the scoped integrity vector (106); the recording of the removed address, the subsisting condition, and the

window; the rule that the condition is not lifted by the removal, so it is priced rather than given effect; and the self-contained computation of all three predicates from the agent's own append-only lineage field (104).

The author's fuller exposition of the same invention carries additional detail that this paragraph does not, and this article does not assert it as filed. Specific numeric parameter values used in that exposition to illustrate the statistic, such as an example recurrence bound or example epoch-count intervals, are not part of the filed paragraph and are not stated here. The fuller exposition also routes the charge through the refusal-counter deflection procedure, whereas the filed paragraph applies a declared modification to the scoped integrity vector; this article follows the filed paragraph. The fuller exposition further locates the declared window in the agent's own signed policy object, a specificity the filed paragraph leaves as an unqualified declared window. Alternative embodiments adjacent to this mechanism in the fuller exposition, including a plurality of named governance addresses, carry-across of the statistic on substitution of the principal, transmission of an escalation record to the governance address, per-scope-partition computation of the recurrence count, and interval weighting of the count, are not part of this paragraph and are outside the scope of this article. Nothing here is a representation of patent coverage; the scope of any resulting patent right will be defined by the claims of a nonprovisional application, should one be filed.

Assertion-Cost Symmetry (/assertion-cost-symmetry)

All 49 steps → (/inventive-steps)

Making an accusation costs the accuser too — symmetric, metered, replay-proof.

Provisional application

PRIMARY TECHNICAL DISCLOSURE

– Assertion-Cost Symmetry: Making an Accusation Cost the Accuser (/articles/assertion-cost-symmetry/making-an-accusation-cost-the-accuser)

SECONDARY TECHNICAL

- Withheld Delegating Faculty, Preserved Direct Execution (/articles/assertion-cost-symmetry/withheld-delegating-faculty)
- Depth-Propagated Harm Inheritance Across a Delegation Chain (/articles/assertion-cost-symmetry/depth-propagated-harm-inheritance)
- Co-Signature Conditioning of Restoration Authority (/articles/assertion-cost-symmetry/co-signature-conditioned-restoration)
- **The Anti-Evasion Charge on Governance-Address Removal (/articles/assertion-cost-symmetry/anti-evasion-governance-address-removal)**
- The Progressive Issuance Decrement (/articles/assertion-cost-symmetry/progressive-issuance-decrement)

TERMINOLOGY

- acknowledgment artifact (/glossary#acknowledgment-artifact)
- assertion-cost counter (/glossary#assertion-cost-counter)
- authorization budget (/glossary#authorization-budget)
- dynamic agent hash chain (/glossary#dynamic-agent-hash-chain)

Assertion-Cost Symmetry overview → (/assertion-cost-symmetry)