Nick Clark · Published July 23, 2026 · PDF

Attack-Conditional Propagation Suspension

A persistent semantic agent that partitions its governance state by context gains isolation, but the couplings that let one context inform another also open an attack surface: flood one context with conduct evaluation artifacts and the resulting state movement can reach a context where the attacker holds no position. Attack-conditional propagation suspension, disclosed in U.S. Provisional Application No. 64/117,812, contains that flood without disabling the agent. The agent computes a per-partition receipt rate over a declared window and an aggregate rate across the remaining partitions taken as the greatest of their rates. When one partition's rate exceeds a declared bound while the aggregate stays below it, the agent suspends cross-partition propagation out of that partition alone, keeps acting within it, and appends a containment determination, without ever writing the authorization gate.

1. The Gap

Partition a semantic agent's governance state by context and you gain genuine isolation. Conduct recorded in one context modifies the scoped integrity vector (106), increments the refusal counter (304), and can write the authorization gate (300) for that context alone, leaving the agent free to act everywhere else. A signed policy object (112) can then declare a coupling so that movement in one partition propagates into a related

one, applying a fraction of the source magnitude to the second partition's scoped integrity vector (106). That is what lets a determination in one context bear on an adjacent one.

The coupling is also a lever. A party that can get conduct evaluation artifacts (116) admitted to a context where it holds a position, but not to a coupled context where it holds none, can flood the first context. Each admitted artifact drives state movement, and the declared coupling carries a fraction of that movement into the second context. By volume alone, an adversary reaches a partition it was never admitted to, moving that partition's scoped integrity vector (106) through the coupling it never had to satisfy.

A response that answered the flood by writing the authorization gate (300) to the withheld state (310) would stop the agent from acting, converting the flood into an outage anyone could trigger on volume. So the requirement is sharper than it first looks. The response has to cut the reach the flood is exploiting while leaving the agent's capacity to act untouched, and it has to separate a localized flood from ordinary load that is merely high across the board. That is the gap this mechanism fills.

2. Mechanism

The agent measures arrival, not content. Over a window declared in the signed policy object (112), it computes for each scope partition a **partition receipt rate**: a count of conduct evaluation artifacts (116) received bearing that partition's identifier. It then forms an **aggregate rate** across the remaining partitions, taken as the greatest of their individual partition rates. The aggregate is the maximum of the others rather than their sum, so it stays below a given bound exactly when every remaining partition individually stays below it. Taking the maximum is what lets the agent separate a spike concentrated in one context from load that is high everywhere.

The suspension condition follows directly. Responsive to a first partition's rate exceeding a declared bound while the aggregate across the remaining partitions stays below that same bound, the agent acts. One context is anomalously hot; the rest are quiet. That is the signature of a targeted flood rather than a busy day, and the agent responds with three moves.

First, it suspends cross-partition propagation out of the first partition only. Movement recorded on conduct in that context stops flowing through the declared coupling into any other partition. The reach the attacker was buying with volume is cut off at its source.

Second, it continues applying its movements within that partition. The flood is not ignored inside the context where it lands: the partition's own scoped integrity vector (106), self-esteem aggregate (108), and refusal counter (304) keep moving under the ordinary rules. Containment withdraws the partition's outward reach, not the consequences that belong to it locally.

Third, it appends a **containment determination** recording the partition, both rates, the bound applied, and the window over which the rates were computed. The event is auditable afterward against the exact figures that produced it.

The load-bearing invariant sits across all three moves: the authorization gate (300) is not written by reason of suspension. Containment removes reach into other partitions, never the agent's ability to act. The agent keeps executing in the flooded partition and in every other one whose gate is granting. Because the response never touches capability, a flood cannot be turned into a shutdown. Suspension is temporary and self-clearing: propagation resumes after the rate stays below the bound for a declared interval.

Two refinements extend the same measurement to close two evasions, and both are part of the filed disclosure.

The first answers an attacker who spreads the flood evenly so that no single partition stands out against the others. Precisely where the per-partition condition fails because the aggregate across the remaining partitions also exceeds the bound, the agent evaluates a **global saturation** condition. It computes a global receipt rate as the sum of every partition's rate over the window, and, responsive to that sum exceeding a global bound declared in the signed policy object (112), suspends cross-partition propagation out of every partition and appends a distinct global containment determination identifying every partition, each partition rate, the global rate, and the bound. The gate is again not written. Distributing the attack no longer buys immunity; it simply trips the global remedy instead of the per-partition one.

The second refinement prices identity into the trigger. The per-partition suspension bound is multiplied by a factor of one less the proportion of that partition's conduct evaluation artifacts (116) that originate from origin-equivalence classes (200) designated untested. A partition whose artifacts all originate from untested classes trips containment at a rate approaching zero; one where none do trips at the full declared bound. Saturation mounted from freshly minted identities is therefore cheaper to contain than saturation from counterparties that hold recorded histories: spinning up new identities to carry the flood lowers the bound at which the flood is contained rather than raising the cost of containing it.

3. Operating Parameters

Every threshold this mechanism turns on is declared in the signed policy object (112), and the filed specification fixes no numeric values for them. The operator, not the disclosure, sets the policy.

The declared parameters are the window over which the partition receipt rates are computed; the per-partition bound the receipt rate is tested against; the interval for which the rate must stay below the bound before propagation resumes; and, for the global saturation remedy, the global bound the summed rate is tested against. The

aggregate rate is not itself a declared value: it is derived at evaluation time as the greatest of the remaining partitions' rates and compared against the same per-partition bound.

The untested-origin weighting is likewise derived rather than declared as a constant. The effective bound for a partition is the declared bound multiplied by one less the proportion of that partition's artifacts originating from origin-equivalence classes (200) designated untested, so the weighting moves with the composition of traffic actually observed. The mechanism assumes a plurality of scope partitions, that being the setting in which cross-partition propagation and its suspension have meaning at all.

4. Composition

This suspension is the third layer of a scope-partitioned containment structure and does not stand alone. Beneath it, the confinement layer maintains, per partition, an independent instance of the scoped integrity vector (106), the self-esteem aggregate (108), the refusal counter (304), the authorization budget (404), and a segment of the append-only lineage field (104), keyed by a partition identifier the conduct descriptor references. Above that, the coupling layer declares, per destination partition and per empathy-scope designation, whether and at what magnitude movement propagates, so a partition whose coupling declaration enumerates nothing is already fully confined. Attack-conditional suspension governs the couplings that layer declares, switching an affirmatively declared propagation off under load and back on when the load subsides.

It composes tightly with the two remedies described above. The global saturation condition is evaluated at exactly the point where the per-partition condition declines to fire, and the untested-origin weighting reshapes the same bound the per-partition condition tests, so the three read as one graduated response to one family of attacks rather than as three unrelated checks.

The gate-not-written invariant is what keeps this mechanism separate from the rest of the filing's control surface. Elsewhere in the architecture the authorization gate (300) is written to the withheld state (310) as the consequence of adjudicated conduct, of deviation, and of assertion cost. Containment is a different kind of constraint: it acts on propagation between partitions, leaves the gate alone, and keeps applying the flooded partition's own movements locally, including its refusal counter (304), under the ordinary rules. A reader who conflates the two would expect containment to stop the agent; it never does.

5. Prior-Art Distinction

Rate limiting and volumetric denial-of-service mitigation detect abnormal request volume and respond by throttling, dropping, or blocking the offending source. The structural difference here is what the response touches. This mechanism neither drops the flooding artifacts nor blocks their source; it admits and prices them within the flooded partition and suspends only their propagation outward. The agent's capacity to act is never the lever, so the flood cannot be converted into an outage.

Circuit-breaker and bulkhead patterns in distributed systems isolate a failing dependency to keep a fault from cascading. They share the instinct to partition and to cut a link under stress, but they trip on the error or latency of a downstream call and they open a path so that callers fail fast. Here the trigger is a receipt-rate anomaly of governance artifacts measured against a signed-policy bound, the thing cut is state-movement coupling between governance contexts rather than a service call, the localized-versus-distributed distinction is drawn explicitly by comparing one partition against the maximum of the others, and every trip is recorded as a containment determination with the figures that caused it.

Sybil-resistant reputation systems down-weight the influence of new or unvouched identities. The untested-origin weighting rhymes with that goal but applies it to a containment threshold rather than to a reputation score: the proportion of untested-

class traffic lowers the bound at which propagation is contained, making a flood of fresh identities cheaper to stop, without computing any standing for the parties involved.

Multi-tenant isolation and namespace separation keep tenants from reaching each other's resources by static configuration. This mechanism isolates the propagation of governance state, not access to resources, and it does so conditionally: coupling is affirmatively declared and permitted in the normal case, and suspension is an attack-conditional response that reverses itself when arrivals return to normal.

None of these comparisons asserts that any product or party practices the claimed mechanism. They mark the boundaries of adjacent categories so the contribution is legible against them.

6. Disclosure Scope

The operative disclosure is U.S. Provisional Application No. 64/117,812. Attack-conditional propagation suspension is disclosed at Section 10.2, paragraph [0355], with its supporting structure at paragraphs [0353] and [0354] and its two extending remedies at paragraphs [0356] (global saturation-defeat) and [0357] (untested-origin-weighted bound).

Disclosed there are the per-partition receipt rate over a declared window; the aggregate rate taken as the greatest of the remaining partitions' rates; the suspension condition of one partition exceeding the bound while the aggregate stays below it; suspension of cross-partition propagation out of that partition alone with continued movement within it; the containment determination recording the partition, both rates, the bound, and the window; the invariant that the authorization gate is not written by reason of suspension; resumption after the rate stays below the bound for a declared interval; the global saturation remedy; and the untested-origin weighting of the bound.

This article states no numeric threshold, window, or interval, because the filing declares each as a parameter of the signed policy object and fixes no value. Nothing here should be read to disclose material beyond the filed provisional. Specific parameter-selection rules, alternative rate constructions, and further partition-handling variants that appear only in the applicant's fuller internal exposition are outside the scope disclosed by this publication and are reserved.

Integrity Quantities and Permitted Deviation (/permitted-deviation)

All 49 steps → (/inventive-steps)

Bounded, accountable deviation from declared values — measured, not forbidden.

Provisional application

PRIMARY TECHNICAL DISCLOSURE

- Permitted Deviation: Bounded, Recorded Departure From an Agent's Declared Values (/articles/permitted-deviation/bounded-departure-from-declared-values)

SECONDARY TECHNICAL

- A Typed, Dimensionless Conduct Divergence Measure (/articles/permitted-deviation/typed-dimensionless-divergence-measure)
- The Policy-Difference Set and Policy-Explained Divergence (/articles/permitted-deviation/policy-explained-divergence)
- Residual Divergence and Its Confined Consumption (/articles/permitted-deviation/residual-divergence-confined-consumption)
- Propagation Halt on Undischarged Reparation Arcs (/articles/permitted-deviation/propagation-halt-undischarged-arcs)
- **Attack-Conditional Propagation Suspension (/articles/permitted-deviation/attack-conditional-propagation-suspension)**

TERMINOLOGY

- aggregate retention (/glossary#aggregate-retention)

- deviation deductible (/glossary#deviation-deductible)
- deviation likelihood (/glossary#deviation-likelihood)
- entropy-weighted harm coefficient (/glossary#entropy-weighted-harm-coefficient)
- need quantity (/glossary#need-quantity)
- permitted deviation (/glossary#permitted-deviation)
- self-esteem aggregate (/glossary#self-esteem-aggregate)

Integrity Quantities and Permitted Deviation overview → (/permitted-deviation)