# VAC PROTOCOL

## Verified Authority Chain

Biometric Human-to-Agent Attribution

for AI Agent Systems

Technical Whitepaper v4.0

February 2026

vacprotocol.org

Violet Shores Pty Ltd

ACN 154 978 122 | ABN 50 154 978 122

## Table of Contents

# Abstract

The Verified Authority Chain (VAC) Protocol addresses a fundamental gap in AI agent security: the inability to prove, at any point in an agent's operation, that a specific verified human authorised the action being taken. Current agent security frameworks secure the agent — its credentials, permissions, and deployment environment — but do not maintain a verifiable link between agent actions and the human who directed them. This paper presents the VAC Protocol's complete architecture: multi-modal biometric human verification, Verified Authority Token (VAT) propagation through multi-agent chains, organisational trust hierarchies, multi-party biometric authorisation, and a graduated maturity model for assessing human-agent attribution security. The protocol is designed as an interoperable standard that complements existing security frameworks (NIST SP 800-63, 800-207, AI RMF) by adding the missing biometric human attribution layer.

*Keywords:* AI agent security, biometric verification, human-agent attribution, verified authority token, multi-agent trust propagation, agent delegation, non-repudiation, zero trust, agentic AI

# 1. The Attribution Gap

AI agent systems are capable of planning and taking autonomous actions that impact real-world systems. They manage financial transactions, access sensitive data, modify production infrastructure, and interact with other agents and humans on behalf of their operators. The security implications are profound.

The current security paradigm for AI agents focuses on three areas: securing the model (robustness to prompt injection, data poisoning), securing the scaffold (least privilege, sandboxing, monitoring), and securing the deployment environment (network isolation, access controls). These are necessary. They are not sufficient.

They all miss a fundamental question: **who is the human behind this agent, and are they still present and authorised right now?**

## 1.1 The Problem with Credentials

In virtually all deployed agent systems, agents operate under credentials — API keys, OAuth tokens, service accounts, or equivalent mechanisms. These credentials prove that someone, at some point, was granted access. They do not prove:

- That the person who created the credentials is the person currently directing the agent
- That the person is still employed, authorised, or present
- That the credentials haven't been shared, stolen, or replayed
- That the agent's current actions are consistent with the human's current intent

This is the attribution gap. The agent's actions are attributed to a credential, not to a verified human. The credential is a proxy, and proxies can be compromised.

## 1.2 The Multi-Agent Amplification

The attribution gap compounds in multi-agent systems. When a coordinator agent delegates to a specialist agent, which may delegate to a sub-agent, the link between the root human and the acting agent is severed within 1–2 levels of delegation. By the third agent in a chain:

- There is no mechanism to verify the root human is still present
- There is no mechanism to verify the root human authorised this specific delegation path
- There is no mechanism to constrain the downstream agent to the root human's intended scope
- There is no mechanism to revoke authority if the root human departs or is compromised

As agent task horizons extend (from minutes to hours), delegation chains deepen (3, 5, 10+ levels), and cross-organisational agent interactions emerge, the attribution gap becomes a systemic risk.

## 1.3 Why This Matters Now

Regulatory frameworks are converging on requirements for human accountability in AI agent actions:

| Framework | Requirement | Gap |
|---|---|---|
| EU AI Act | Human oversight for high-risk AI systems | No attribution mechanism specified |
| NIST AI RMF | Accountability and traceability | No biometric binding standard |
| NIST SP 800-207 | Zero trust: never trust, always verify | Verifies agents, not the humans behind them |
| OWASP Agentic Top 10 | Identity and privilege abuse prevention | Credential-based identity only |
| SEC/FINRA | Non-repudiation for financial transactions | Digital signatures, not biometric proof |

The VAC Protocol provides the missing technical mechanism that these frameworks require but do not yet specify.

# 2. Protocol Architecture

The VAC Protocol operates in three layers: Identity, Delegation, and Attribution. Each layer builds on the previous one.

## 2.1 Layer 1: Biometric Identity Verification

At the foundation, VAC verifies the identity of the human directing an agent operation using multi-modal biometric verification. The system combines four independent modalities in a single gesture:

| Modality | Signal | Attack Resistance |
|---|---|---|
| Facial geometry | 3D facial landmarks, depth mapping, micro-expressions | Defeats 2D face swaps; requires real-time 3D deepfake (not yet practical) |
| Voice pattern | Vocal biomarkers, prosody, spectral analysis | Requires simultaneous voice clone synced with facial output |
| Behavioural biometrics | Keystroke dynamics, touch patterns, gait | Requires cloning of unconscious behavioural patterns — no known attack |
| Device context | Hardware attestation, location, network fingerprint | Requires physical access to the specific enrolled device |

The mathematical foundation for why four modalities represent the critical threshold is detailed in the companion paper, *Why Four Modalities? A Mathematical Framework for Multi-Modal Authentication Security* (VAC Technical Whitepaper v1.0). Key findings: with four independent modalities, real-time simultaneous spoofing requires coordinated defeat of all four channels within a session window. The Multi-Modal Attack Resistance Time (MART) metric demonstrates that four-modality verification pushes the minimum attack time beyond practical real-time session windows.

## 2.2 Continuous Trust Scoring

Unlike binary authentication (pass/fail at session start), VAC computes a continuous trust score that evolves throughout the session:

- **Trust score range:** 0.0 (no trust) to 1.0 (maximum confidence)
- **Initial score:** Set at session start based on biometric verification confidence across all four modalities
- **Decay function:** Trust score decays over time since last verification, with decay rate proportional to the sensitivity of active operations
- **Re-verification boost:** Periodic or triggered biometric re-verification restores the trust score

- **Action gating:** Each action type has a minimum trust score threshold. If the current score falls below the threshold, the action is blocked until re-verification

This replaces the binary “authenticated/not-authenticated” model with a continuous signal that more accurately represents the confidence that the verified human is still present and in control.

# 3. Verified Authority Token (VAT)

The Verified Authority Token is the central cryptographic object in the VAC Protocol. It carries the verified human's biometric provenance through arbitrarily deep multi-agent chains.

## 3.1 Token Structure

A VAT is a signed, structured data object containing:

| Component | Contents |
|---|---|
| Header | Token version, signing algorithm, token type (root or derived) |
| Identity | Verified human identity reference (cryptographic hash, not plaintext PII) |
| Trust | Trust score at time of creation, minimum trust threshold for this token |
| Scope | Resource types, action types, data domains, temporal windows, sensitivity thresholds |
| Delegation | Maximum delegation depth, current depth, parent token reference, delegation chain metadata |
| Context | Organisational context reference (if applicable), multi-party authorisation references (if applicable) |
| Validity | Not-before timestamp, not-after timestamp, re-verification requirements |
| Signature | Cryptographic signature binding the payload to the biometric attestation |

## 3.2 Token Lifecycle

### Creation (Root VAT)

When a biometrically-verified human authorises an agent operation, the system generates a root VAT. The root VAT's scope, trust score, and validity are set based on the human's verified identity, organisational role, and the nature of the requested operation.

### Derivation (Delegated VAT)

When a coordinator agent delegates to a specialist agent, it creates a derived VAT from its own VAT. Derivation rules enforce strict narrowing:

- **Scope:** Set intersection of parent scope and requested scope. A derived token can never access resources or perform actions outside its parent's scope.
- **Trust score:** Decreased as a function of parent trust score, delegation depth, and delegating agent's own trust properties. Deeper delegation inherently carries lower trust.
- **Validity:** Minimum of parent expiry and requested expiry. A derived token cannot outlive its parent.

- **Delegation depth:** Incremented by one. Cannot exceed the maximum specified in the root token.

### Verification

Before any agent action is permitted, the receiving system verifies the presented VAT:

- Verify the signature chain from presented token back to root token
- Verify the root token traces to a biometric attestation from a recognised verification provider
- Verify the trust score meets the minimum threshold for the requested action
- Verify the action falls within the scope encoded in the token
- Verify the token has not expired and delegation depth does not exceed maximum
- Check revocation status of root and all intermediate tokens
- Record the verification event in the audit trail

### Revocation

Revocation cascades through the entire token chain:

- **Root revocation:** Human ends session, trust score drops below threshold, or explicit revocation. All derived tokens in all chains are immediately invalidated.
- **Intermediate revocation:** Compromised agent or organisational change. All tokens derived from the revoked token are invalidated.
- **Latency target:** Sub-second for root revocation, under 5 seconds for full chain propagation.

# 4. Organisational Trust Hierarchies

In enterprise deployments, agent authority derives not just from individual identity but from organisational context. The VAC Protocol supports hierarchical delegation with strict narrowing at every level:

**Organisation → Group/Role → Human → Agent → Action**

## 4.1 Hierarchical Delegation

An organisation registers as a trust root in the VAC system with verifiable external identifiers (ABN, LEI, DUNS, or equivalent). The organisation defines roles and groups, each with specific authority scopes. Humans are verified biometrically and bound to their organisational role(s). Agents inherit authority from the human's role-scoped authority, which is further narrowed by the task scope.

At every level, authority can only narrow. An organisation's compliance department cannot grant an agent broader access than the compliance role permits. A human within that role cannot grant an agent broader access than their individual authority permits.

## 4.2 Revocation Propagation

Organisational changes propagate instantly through all active agent chains:

- **Employee departure:** All agents operating under the departed employee's organisational delegation are immediately suspended.
- **Role change:** Agent permissions automatically adjust to the new role's authority scope.
- **Organisational sanction:** Regulatory action against an organisation suspends all agents operating under any human within that organisation.

## 4.3 Multi-Organisational Identity

A single verified human may hold roles in multiple organisations. The VAC Protocol supports context switching between organisational identities, with biometric re-verification required at each switch. Conflict-of-interest detection operates across organisational contexts, flagging when agents from a human's different organisational roles interact.

## 4.4 Individual Trust Score and Role Authority Interaction

An agent's effective authority is always the intersection of two independent constraints: the individual human's current biometric trust score, and the organisational role's granted permissions. Neither can compensate for the other.

If a human's trust score drops — due to failed re-verification, anomalous behavioural signals, or an elapsed re-verification window — all agents operating under that human's authority are immediately constrained, even if the organisational role technically permits the actions being performed. Real-time biometric signals reflecting the human’s actual presence always take precedence over static role assignments.

Conversely, a high individual trust score cannot override an absent role permission. A human with perfect biometric confidence but no procurement authority cannot authorise procurement agents, regardless of their trust score.

This minimum-of-both rule applies at every level in the delegation chain, ensuring the system never relies on a single authority source.

# 5. Multi-Party Biometric Authorisation

High-stakes agent operations may require biometric verification from multiple designated humans before execution. The VAC Protocol implements M-of-N biometric authorisation — multi-signature with proof of life, not proof of credential.

## 5.1 Patterns

| Pattern | Description | Example |
|---|---|---|
| Threshold (M-of-N) | Any M of N designated humans must biometrically verify before agent proceeds | 3 of 5 board members approve a major transaction |
| Sequential chain | Approvals must occur in a defined order, each biometrically verified | Analyst → manager → director approval for a trade |
| Role-combined | Approvals required from humans in different organisational roles | Both physician and privacy officer approve a data release |
| Biometric veto | Any designated human can biometrically verify to halt an agent operation | Compliance officer vetoes a risky automated trade |

## 5.2 Properties

- Each approval requires the designated individual’s biometric presence — no proxy approvals
- All approvals must be collected within a configured time window
- A composite VAT is generated incorporating the verified identities of all approving humans
- The composite VAT propagates through the agent chain with the combined authority
- Re-verification triggers for high-sensitivity actions within the chain require re-verification from the original M-of-N threshold

# 6. Attribution Maturity Model

The VAC Protocol defines a five-level maturity model for assessing the strength of human-to-agent attribution in any AI agent deployment. The model enables standardised assessment across platforms, deployments, and regulatory contexts.

| Level | Name | Characteristic | Attribution Strength |
|---|---|---|---|
| 1 | None | Agent operates under static credentials; no human attribution | Anyone with credentials can direct the agent; full repudiation possible |
| 2 | Credential-Bound | Agent actions logged with the credential that authorised them | Attribution to a credential, not a person; shared credentials break attribution |
| 3 | Session-Verified | Human authenticates at session start; actions attributed to session | No proof of continued presence; vulnerable to session hijacking |
| 4 | Biometrically Verified | Biometric verification at session start; actions cryptographically attributed | Proof of presence at initiation only; single-agent systems |
| 5 | Continuously Verified (VAC) | Biometric verification with re-verification triggers; VAT propagation through chains | Non-repudiation for every action at any delegation depth |

**Most deployed agent systems today operate at Level 1 or Level 2.** Level 3 is emerging in enterprise deployments with SSO integration. Level 4 is achievable with current biometric technology but requires integration not yet standard in agent platforms. Level 5 requires the full VAC Protocol architecture.

## 6.1 Applications

- **Regulatory compliance:** Regulators can specify minimum attribution levels for different categories of agent operations (e.g., financial agents must achieve Level 4+)
- **Procurement criteria:** Enterprises can require agent platforms to support specific maturity levels
- **Counterparty risk:** Organisations can assess the attribution risk of interacting with another organisation’s agents based on their maturity level
- **Insurance underwriting:** Cyber insurers can price policies based on attribution maturity, as higher maturity reduces fraud and repudiation risk

# 7. Conformance Testing and Performance Metrics

The VAC Protocol introduces security properties — biometric human-agent attribution, trust propagation through delegation chains, scope narrowing enforcement, and cascading revocation — that have no existing industry testing standards. MART and CFAR address biometric verification performance (Layer 1), but do not cover the trust propagation, delegation, revocation, and conformance properties introduced in Layers 2 and 3.

Without a conformance testing framework, implementers cannot verify that their deployment actually delivers the security guarantees it claims. A system that passes biometric verification (MART/CFAR) but fails to enforce scope narrowing in delegation chains provides a false sense of security.

This section defines standardised metrics, conformance tests, and performance benchmarks for evaluating a complete VAC Protocol implementation.

## 7.1 Novel Metrics

### VAT Verification Time (VATV)

The end-to-end time required to verify a complete VAT chain from the presented token back to the biometric attestation at the root:

```
VATV(d) = T_verify(token) + T_chain_walk(d) + T_root_check +
T_revocation_check
```

Where d is the delegation depth. VATV must scale sub-linearly with depth for the protocol to be practical in deep chains. Reported at delegation depths of 1, 3, 5, 10, and 20 (median and 99th percentile).

### Revocation Propagation Latency (RPL)

The time from a revocation event to the last affected agent in the chain being notified and suspended:

```
RPL = T_last_agent_suspended − T_revocation_initiated
```

Measured separately for three revocation types:

- **Root revocation** (human-initiated): target < 1 second
- **Intermediate revocation** (agent compromise): target < 3 seconds
- **Organisational revocation** (role change, departure): target < 5 seconds
- **Full chain propagation** at depth d: target < 5 seconds for d ≤ 20

### Scope Narrowing Enforcement Rate (SNER)

A conformance metric measuring the percentage of delegation events where scope narrowing is correctly enforced:

```
SNER = (derived_scope ⊆ parent_scope delegations) / (total delegations) × 100
```

SNER must equal 100.0% for a conformant implementation. Any value below 100% indicates a scope escalation vulnerability. This is a pass/fail conformance test, not a performance benchmark.

### Trust Score Propagation Accuracy (TSPA)

Measures the accuracy of trust score computation through delegation chains:

```
TSPA = |computed_trust − expected_trust| / expected_trust
```

Where expected_trust is derived deterministically from the root trust score, decay function, delegation depth, and delegating agent properties. TSPA must be < 0.001 (0.1% error) for conformance.

### Attribution Maturity Level Assessment Score (AMLAS)

A structured assessment that evaluates an agent system against the five-level maturity model (Section 6), producing:

- An assigned maturity level (1–5) based on verifiable capabilities
- A capability coverage percentage within the assigned level
- A gap analysis identifying capabilities missing for the next level
- A counterparty risk rating based on the assessed level

## 7.2 Conformance Test Suite

A conformant VAC Protocol implementation must pass all six test categories:

### Category 1: Token Structural Conformance

Validates that VATs contain all required fields (header, identity, trust, scope, delegation, context, validity, signature), that signatures are cryptographically valid and chain to a recognised biometric attestation, and that extension fields do not conflict with core fields.

### Category 2: Derivation Rule Conformance

Validates that derived scope is provably a subset of parent scope (set intersection), trust scores are provably ≤ parent adjusted by the specified decay function, validity periods are provably ≤ parent, delegation depth increments by exactly one per derivation, and maximum delegation depth is never exceeded.

### Category 3: Revocation Conformance

Validates that root revocation invalidates all derived tokens within RPL target, intermediate revocation invalidates downstream without affecting upstream, revoked tokens are rejected on subsequent verification, and partial chain revocation does not affect unrelated chains.

### Category 4: Multi-Party Authorisation Conformance

Validates that M-of-N thresholds are correctly enforced (M−1 approvals do not authorise), each approval is linked to a distinct biometrically-verified human, approvals are collected within the configured time window, expired partial approvals do not carry forward, and composite VATs correctly encode all approving identities.

### Category 5: Organisational Hierarchy Conformance

Validates that authority flows strictly downward (organisation → group → human → agent), cross-context agent permissions are scoped to the active organisational context, organisational revocation propagates to all humans and agents within scope, and multi-organisational identity binding correctly isolates contexts.

### Category 6: Cross-Platform Interoperability Conformance

Validates that VATs generated by one platform are verifiable by another implementing the same specification, lightweight verification (without full VAC implementation) correctly validates token chains, extension fields from one domain do not break verification on platforms that do not implement that domain, and API endpoints for create, derive, verify, and revoke conform to the specification.

## 7.3 Performance Benchmark Suite

A standardised benchmark suite produces a comparable performance profile for any VAC Protocol implementation:

- **Verification throughput:** VATV at delegation depths 1, 3, 5, 10, 20 (median and 99th percentile)
- **Revocation performance:** RPL for root, intermediate, and organisational revocation (median and 99th percentile)
- **Creation throughput:** Maximum sustained VAT creation rate (tokens per second)
- **Verification rate:** Maximum sustained VAT verification rate (verifications per second)
- **Resource footprint:** Memory footprint per active token chain; network bandwidth per verification event

The benchmark enables procurement teams, regulators, and certification bodies to compare implementations objectively across a standardised set of performance dimensions.

## 8. Standards Alignment

The VAC Protocol is designed to complement, not replace, existing security frameworks. It provides the biometric human attribution layer that these frameworks reference but do not specify:

| Standard | Current Scope | VAC Extension |
|---|---|---|
| NIST SP 800-63-4 | Digital identity; password/token/MFA authentication | Multi-modal biometric verification; continuous trust scoring replaces binary auth |
| NIST SP 800-207 | Zero trust: never trust, always verify | Verify the human behind the agent at every action point via VAT |
| NIST AI RMF (100-1) | Accountability and traceability for AI systems | Cryptographic mechanism: every agent action traceable to a verified human |
| NIST AI 600-1 | GenAI risk profile; information security considerations | Biometric attribution addresses GenAI-specific identity and non-repudiation risks |
| OWASP Agentic Top 10 | Identity abuse; tool misuse; cascading failures | Biometric binding prevents identity abuse; trust narrowing limits cascading scope |
| EU AI Act | Human oversight for high-risk AI | Verifiable human oversight: biometric proof of human presence and authorisation |
| ISO/IEC 27001 | Information security management systems | Verified Contribution Ledger: legally admissible audit trails with biometric non-repudiation |
| FIDO2/WebAuthn | Passwordless authentication; device-bound credentials | Extends beyond device binding to continuous biometric human presence verification |

# 9. Application to Regulatory Identity Verification (KYC/AML)

The VAC Protocol's biometric human verification has a direct application beyond agent security: regulatory identity verification (Know Your Customer / KYC) for financial services. As AI agents increasingly operate financial accounts on behalf of verified humans, the attribution gap becomes a regulatory compliance gap.

## 9.1 The KYC Problem

Current KYC systems verify identity at onboarding (point-in-time) and rely on periodic review. Between reviews, there is no assurance the person conducting transactions is the verified individual. When AI agents operate accounts, this gap widens: the KYC obligation attaches to the human, but actions are performed by agents with no identity binding to the verified human.

Australia's AML/CTF reforms (effective 31 March 2026) are moving KYC from point-in-time verification toward continuous customer due diligence — a shift architecturally aligned with VAC's continuous trust monitoring.

## 9.2 How VAC Addresses KYC

The VAC Protocol extends to KYC through six capabilities:

- **Document-to-biometric cross-reference.** Government-issued identity documents are captured and cross-referenced against live multi-modal biometric verification, generating a document-to-live matching confidence score incorporated into the composite trust score.
- **Human-vouched identity verification.** Real humans who know the individual cryptographically confirm their identity. This provides a novel second verification source independent of document verification — satisfying the regulatory requirement for verification from "at least two separate data sources."
- **Continuous KYC through trust score monitoring.** Rather than periodic review cycles, the trust score is continuously computed. When it falls below a configurable threshold, re-verification is automatically triggered — implementing risk-based enhanced due diligence.
- **Agent-inclusive KYC delegation.** When a verified individual authorises an AI agent to operate a financial account, the system creates a KYC delegation record binding agent actions to the individual's KYC identity. If the trust score drops, all agents are immediately suspended.
- **Presentation attack detection.** Multi-layer liveness verification (passive analysis, active challenges, cross-modal consistency, injection attack detection) conforming to ISO/IEC 30107-3 operates as a prerequisite gate for biometric KYC.
- **PEP/sanctions screening integration.** Biometrically-anchored screening provides higher-confidence matches than name-only screening, with continuous re-screening and automatic trust score adjustment.

## 9.3 Portable KYC Credentials

Upon successful KYC verification, the system can issue a portable KYC credential — a cryptographic attestation extending the Verified Authority Token — that attests to the individual's KYC status without revealing underlying personal data. This credential can be presented to other reporting entities, eliminating repeated verification while maintaining regulatory compliance through the trust graph.

## 9.4 KYC Compliance Scoring

The protocol computes a KYC Compliance Score for each verified individual, aggregating initial verification completeness, continuous monitoring status, agent delegation governance, and evidence retention completeness. This provides a quantitative, auditable, real-time measure of regulatory compliance that can be reported to regulators including AUSTRAC, FinCEN, and equivalent bodies.

# 10. Intellectual Property

The VAC Protocol is protected by the following patent filings:

| Filing | Details |
|---|---|
| Provisional Patent | AU 2026901425, filed 21 February 2026. 112 claims covering multi-modal biometric verification, continuous trust scoring, agent delegation, single-gesture authentication. |
| Supplementary #1 | AU 2026901428, filed 22 February 2026. Claims 113–134 covering zero-knowledge proofs, blockchain oracle integration, biometric aging adaptation, agent resource allocation, unified trust graphs. |
| Supplementary #2 | AU 2026901474, filed 23 February 2026. Claims 135–167 covering verified contributor authority, organisational trust hierarchies, multi-party biometric authorisation, agent chain trust propagation via VAT, attribution maturity model, conformance testing. |
| Supplementary #3 | Claims 168–195 (28 claims). Document-to-biometric KYC, presentation attack detection and liveness verification, PEP/sanctions screening integration, continuous KYC compliance, portable KYC credentials, agent-inclusive KYC delegation. |
| Total coverage | 195 claims across 34 sections covering identity verification, agent delegation, trust propagation, organisational hierarchies, multi-party authorisation, conformance testing, KYC compliance, presentation attack detection, and regulatory screening. |
| Assignee | Violet Shores Pty Ltd (ACN 154 978 122) |
| Priority date | 21 February 2026 |

The protocol is being developed as an open standard for human-to-agent attribution. Licensing terms for implementation will be published at vacprotocol.org.

# 11. Conclusion

AI agent security requires more than securing the agent. It requires securing the link between the agent and the human who directed it.

Existing security frameworks provide the walls, the locks, and the cameras. The VAC Protocol provides the one thing they cannot: proof that a specific verified human authorised this specific agent action at this specific time — through any depth of delegation, across any number of agents, with trust that can only narrow and authority that can be revoked in sub-second timeframes.

As agent task horizons extend, delegation chains deepen, and cross-organisational agent interactions become the norm, the attribution gap will become the defining security challenge of the agentic era. The VAC Protocol addresses this challenge now, before the gap becomes a systemic risk.

Protocol specification and updates: vacprotocol.org

Interactive simulation: vacprotocol.org/simulation.html — 7 deployment scenarios with step-through narrative

**Technical enquiries:** admin@violetshores.com

**Regulatory engagement:** Responding to NIST CAISI RFI on AI Agent Security (NIST-2025-0035) and NCCoE concept paper on Software and AI Agent Identity and Authorization.