VAC PROTOCOL

Verified Authority Chain

Biometric Human-to-Agent Attribution

for AI Agent Systems

Technical Whitepaper v3.0

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.

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. 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

8. 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	Claims 135–162 (28 claims). Verified contributor authority, organisational trust hierarchies, multi-party biometric authorisation, agent chain trust propagation via VAT, attribution maturity model, VAT token specification.
Total coverage	162 claims across identity verification, agent delegation, trust propagation, organisational hierarchies, multi-party authorisation, attribution maturity assessment, and interoperable token specification.
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.

9. 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

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.