The Four Dimensions of Identity Governance

Extending TRUST, CRED, PRIV, and BEHAVE to Agentic AI and the Post-Quantum Era

By Ross Foard

Executive Summary

In November 2016, I presented a framework at CISA that proposed four dimensions every enterprise must continuously evaluate for every principal operating on its networks: TRUST, CRED, PRIV, and BEHAVE. These dimensions became the organizing logic of the CDM program's Identity and Access Management capability, subsequently formalized in the CDM ICAM Reference Architecture v1.3 (September 2023) as the framework evolved through Zero Trust adoption.

Two converging challenges now require that framework to evolve again — simultaneously.

The first is post-quantum cryptography (PQC): the cryptographic substrate beneath all four dimensions — the PKI that underpins TRUST, the credentials that operationalize CRED, the session tokens that enforce PRIV, and the audit signatures that anchor BEHAVE — is vulnerable to quantum attack. The Harvest Now, Decrypt Later (HNDL) threat is not theoretical. Adversaries are collecting encrypted IAM artifacts today, intending to decrypt and forge them when a cryptographically relevant quantum computer (CRQC) exists. OMB Memorandum M-26-15 (June 2026) establishes binding federal deadlines for PQC migration, and GSA has been directed to establish an inter-agency working group specifically to modernize Federal Identity, Credential, and Access Management (FICAM) to support PQC.

The second is agentic AI: autonomous systems that perceive context, form plans, invoke tools, take consequential actions, and delegate to other agents. They are arriving in production environments without the governance scaffolding they require, and they break foundational assumptions in every one of the four dimensions.

These two challenges interact in ways that compound each other's risk. Agentic AI's reliance on ephemeral, dynamically issued credentials creates new quantum attack surfaces. Multi-agent delegation chains — trust assertions passed from orchestrator to sub-agent at runtime — are exactly the kind of high-value cryptographic artifact an adversary would harvest today for future forgery. And the behavioral opacity that makes agents difficult to govern also makes PQC implementation flaws in agentic pipelines difficult to detect.

This paper makes a unified argument: the four dimensions are necessary and sufficient for governing both post-quantum identity risk and agentic AI identity — but each dimension requires a first-principles evolution that addresses both challenges together. We do not need a fifth dimension. We need to think harder about the four we have, and we need to do so with full awareness that the cryptographic ground beneath them is shifting.

1. Introduction: The Identity Control Plane Under Simultaneous Pressure

Identity governance has always been the enterprise's most foundational security capability. Every access decision — every authentication, every authorization, every federation assertion — flows through the identity control plane. When that control plane is compromised, every downstream security investment is undermined.

For three decades, the identity control plane rested on two assumptions so deeply embedded that they were rarely stated explicitly. First: the principals we govern are human beings, or systems that behave like simplified humans — predictable, persistent, with stable behavioral patterns and pre-assigned privilege. Second: the cryptography beneath our identity assertions is computationally hard to break — RSA keys, elliptic curve signatures, and TLS key establishment are secure enough that we can build long-lived trust relationships on them.

Both assumptions are now under simultaneous pressure — and federal policy has made the urgency explicit.

The post-quantum imperative is now law. NIST finalized three post-quantum cryptography standards in August 2024 (FIPS 203, 204, and 205), establishing ML-KEM for key encapsulation, ML-DSA for digital

signatures, and SLH-DSA as a hash-based signature alternative. Executive Order 14412 (June 2026), Securing the Nation Against Advanced Cryptographic Attacks, mandates an accelerated government-wide migration to PQC with binding deadlines for high-value assets. OMB Memorandum M-26-15 (June 2026) operationalizes those deadlines across five migration phases. Critically for identity professionals, M-26-15 directs GSA to establish an inter-agency working group on modernizing FICAM to support PQC — recognizing that identity infrastructure is not merely one of many systems requiring migration, but the control plane through which all other PQC migration decisions flow. CISA has formally classified ICAM as transitioning in its Product Categories advisory — acknowledging that while key establishment is progressing, the digital signature layer that authenticates every identity assertion remains classically vulnerable. The standards exist. The policy mandates exist. What the field lacks is IAM-specific architectural guidance that addresses both the human identity infrastructure and the novel principal types that the standards were not written to address.

The agentic AI challenge has no policy mandate yet, but it can't wait for one. Agentic AI systems violate the first assumption completely. They are principals that reason, plan, and act with degrees of autonomy that make them qualitatively different from every identity type governance frameworks have addressed before. The same CDM framework that I helped build at CISA, which successfully extended from human identities to devices, service accounts, and non-person entities, has not yet reckoned with what it means to govern a principal that acquires privilege at runtime, delegates to sub-agents mid-task, and can be redirected by adversarial content without any credential compromise. These systems are in production today, without the governance scaffolding they require.

ICAM modernization requires including both of these workstreams into the modernization plan. Treating them as separate programs will produce gaps at precisely the points where they intersect — and as Section 3.3 demonstrates, those intersections are the highest-risk surfaces in the emerging identity landscape.

The TRUST/CRED/PRIV/BEHAVE framework was designed to be durable precisely because it was built around questions, not technologies. Who are you, and have you been vouched for? Can you prove it? What are you permitted to do? Are you acting within your authorized purpose? Those questions apply to every principal type and survive every technology generation. But the answers — the mechanisms, the artifacts, the operational controls — must evolve for each new context.

This paper provides that evolution for the dual context of post-quantum cryptography and agentic AI.

2. The Four Dimensions: A Framing

Before evolving each dimension, it is worth being precise about what the framework is and is not.

The four dimensions are not a checklist. They are not a maturity model. They are not a compliance framework, though they map cleanly to compliance requirements. They are continuous assertions that the enterprise must be able to make about every principal, at every moment, across the full identity lifecycle.

This framework draws on the concept of vectors of trust established in IETF RFC 8485, which defines component dimensions for describing the trust characteristics of identity transactions between parties. Our framing extends that lineage from transaction-level signaling to continuous principal-level governance — a broader and more persistent conception of trust that spans the full identity lifecycle rather than a single authentication event.

TRUST: The provenance and accountability dimension. Every principal must have an identity that can be trusted: each user, each device, each service account, each AI agent. Has this principal been properly vouched for, and does that accountability chain remain intact? TRUST isn't established at onboarding and then assumed valid forever. It is strongly established, and then continuously evaluated to confirm the chain connecting a principal to a real-world authority (a human or organizational sponsor, an issuing organization, an authoritative identity source) remains unbroken and valid.

CRED: The proof-of-claim dimension. Once TRUST is established it must immediately be issued a proof of claim, cryptographically. Can this principal demonstrate that it is who it claims to be? Credentials are the mechanism by which identity assertions become verifiable. What matters is not merely that a credential was issued, but that it was issued by a trusted authority, remains currently valid, is bound to exactly one principal,

is being presented by the principal it was issued to, and that the cryptographic assumptions underlying that binding remain computationally sound.

PRIV: The authorized scope dimension. What is this principal permitted to do, and is that permission still appropriate given current context? Privilege is not a property of the principal alone; it is a relationship between the principal, the resource, the context, and time. The PRIV dimension requires both that privilege be granted with minimum necessary scope and that it be continuously re-evaluated against operational need.

BEHAVE: The runtime conformance dimension. Is this principal acting in a manner consistent with its authorized purpose and established norms? BEHAVE is where identity governance meets detection and response. It is the dimension that catches the legitimately credentialed principal who has been compromised, the authorized user who is misusing access, and the system whose cryptographic implementation has been exploited in ways no credential check would reveal.

Together, these four dimensions define a coordinate space. Every principal exists at a position across all four axes simultaneously. Governing identity means continuously characterizing that position and responding when it moves in ways that indicate risk. A dimensional framing — rather than a checklist framing — matters because it implies that security posture is dynamic: a principal's TRUST position can degrade, credentials expire and can be compromised, privilege can become excessive, and behavior can shift. The enterprise must be able to detect and respond to movement across any dimension, at any time, for any principal.

3. Why Post-Quantum Risk and Agentic AI Require Simultaneous Treatment

It would be tempting to treat post-quantum cryptography migration and agentic AI governance as separate workstreams — one a cryptographic engineering problem, the other a governance problem. That framing is incorrect, and acting on these workstreams independently will produce gaps at exactly the points where the two challenges intersect.

3.1 Post-Quantum Risk: A Cryptographic Foundation Under Known Threat

The quantum threat to IAM manifests in two distinct failure modes with different timelines and migration priorities.

The confidentiality failure mode, Harvest Now, Decrypt Later, is active today. Adversaries are collecting encrypted IAM exchanges: authorization codes, token acquisitions, SAML assertions, SCIM provisioning payloads, privileged API traffic, and federation session parameters. When a CRQC becomes available, Shor's algorithm enables polynomial-time factorization of RSA keys and solution of elliptic curve discrete logarithms, decrypting this harvested material retroactively.

The authenticity failure mode, future signature forgery, is a later but more consequential threat. Digital signatures underpin every trust assertion in the identity control plane: JWT signatures that authenticate tokens, SAML assertion signatures that enable federation, X.509 certificate signatures that establish PKI trust chains, and the signatures on device attestations, firmware measurements, and audit records. A CRQC capable of breaking RSA or ECC signatures would let an adversary forge any of these: impersonating identity providers, manufacturing valid tokens, fabricating compliant device attestations.

Vikram Kalekar's MV-QRIAM framework makes an important architectural point here: IAM is a special case for quantum risk, a mesh of trust boundaries where few authoritative signers (IdPs, CAs, token services) feed many distributed verifiers (resource servers, API gateways, relying parties). An upgrade to quantum-resistant signatures at the signer does not provide quantum resilience until the verifier ecosystem catches up, and that coordination gap, not algorithm availability, is the binding constraint on migration timelines.

CISA's formal classification of ICAM as transitioning, meaning not yet widely PQC-capable, confirms the signature gap. Many IAM products support ML-KEM for key encapsulation, but ML-DSA adoption for digital signatures remains limited. Organizations must not mistake transport-layer PQC progress for ICAM PQC readiness. The authentication layer remains classically vulnerable until signature migration is complete.

3.2 Agentic AI: A Category Break in Principal Governance

Federal agencies and enterprises have extended identity governance frameworks to non-person entities (NPEs): service accounts, API keys, device certificates, RPA bots — with reasonable success. The GSA Digital Worker Identity Playbook anticipated AI-adjacent automation under the NPE umbrella. But agentic AI systems are not NPEs with more sophisticated behavior. They represent a qualitative break from every principal governance model previously developed because they:

- **Have no stable behavioral baseline.** A service account runs the same script every time. An agent reasons about its current context and decides what to do next. The same agent, given a different input, will take a meaningfully different sequence of actions. This makes behavioral anomaly detection structurally inapplicable in its existing form.
- **Acquire privilege dynamically.** Traditional governance works because privilege is pre-assigned: roles modeled, entitlements provisioned, access certified. An agent has a tool manifest and decides at runtime which tools to call, in what order, with what parameters. The privilege surface is a function of the task and the agent's in-context reasoning, not a static entitlement set.
- **Delegate to other agents.** A single agentic task frequently involves an orchestrator spawning sub-agents, each of which may call tools, invoke APIs, write to storage, and communicate externally. The trust and credential questions that are complex for a single agent become exponentially more complex in a multi-agent pipeline.
- **Are opaque at the point of action.** When a human administrator takes a privileged action, there is a cognitive and social accountability chain. When an agent acts, the decision may have emerged from a chain-of-thought reasoning process that is not logged, not inspectable, and not reproducible. The audit trail that every compliance framework requires is broken by default.
- **Can be manipulated mid-task without credential compromise.** Prompt injection attacks — where malicious content in the environment hijacks the agent's reasoning — have no analog in traditional NPE governance. An agent that is legitimately credentialed and authorized can be redirected by an adversary without triggering any existing detection mechanism.

3.3 The Intersection: Why Agent Identity Is Specifically Quantum-Vulnerable

Agentic AI systems create several quantum-specific risks that do not exist for human or traditional NPE identities.

Multi-agent delegation assertions are high-value HNDL targets. When an orchestrator agent grants a sub-agent scoped authority to act on its behalf, that delegation is expressed as a cryptographic assertion — a signed token conveying the scope of delegated trust. These assertions are exactly the kind of artifact a quantum-era adversary would harvest today: they are authoritative, they convey privilege, and they can be forged in the post-CRQC era to impersonate trusted orchestrators. A delegation chain harvested in 2026 and forged in 2031 could inject a rogue sub-agent into a multi-agent pipeline with the full trust of the original orchestrator.

Model provenance attestations have long time horizons. Unlike a human session token that expires in hours, a model provenance attestation — the signed record that a given agent is running model version X with system prompt Y — may be referenced for months or years in audit trails and governance records. If those attestations are signed with classically vulnerable algorithms, they can be retroactively forged, destroying the integrity of the governance record.

Agent audit logs require long-term signature integrity. The evidentiary value of an agent's audit trail depends on the integrity of the signatures protecting those records. An adversary who can forge audit log signatures in the post-CRQC era can rewrite the history of what an agent did — eliminating the forensic foundation that BEHAVE-driven investigation requires.

Ephemeral agents are, paradoxically, better positioned. An agent that receives a short-lived, task-scoped workload identity credential and discards it at task completion has a much smaller HNDL exposure than a human identity with a long-lived session token or a service account with a static API key. The JIT/JEA credential model that agentic AI architectures should adopt for CRED governance aligns naturally with PQC migration: short-lived tokens are lower-value HNDL targets, and short credential lifetimes reduce the window

during which a harvested credential remains actionable even if later decrypted.

4. TRUST: Provenance, Accountability, and Quantum-Resilient Vetting

4.1 Classical TRUST and Its Evolution for Agents

TRUST, in the federal identity model, begins with HSPD-12 and the PIV credential ecosystem. The premise is clear: before a principal receives credentials and access, it has been vetted through suitability determination, identity proofing, and background investigation. NIST SP 800-63A sets the identity proofing requirements that make this vetting step rigorous rather than perfunctory, and SP 800-63 more broadly remains the federal baseline against which any TRUST posture model, including the one this paper proposes, must be evaluated for conformance. TRUST then monitors whether that vetting remains valid through continuous vetting capabilities tied to the Master User Record (MUR).

For NPEs, CDM extended TRUST through human or organizational sponsorship: every non-person entity must have a named accountable sponsor for its existence, configuration, and behavior. That principle extends directly to agents — with additional requirements that reflect agents' unique trust surface.

Sponsor accountability must be a hard deployment gate. Every agent must have a named sponsor before it touches production — a person or organizational role with a traceable accountability chain, who owns the deployment decision, understands the capability scope, and will answer for the agent's actions. The named sponsor must resolve to a human or organizational accountability anchor; another agent cannot serve as a sponsor. In practice today, agents are deployed as product features without identity governance review. That is a TRUST failure before the agent takes its first action.

Model provenance is a TRUST signal. An agent's behavior is shaped by its underlying model: the foundation model, any fine-tuning, the system prompt, and the tool manifest. These are the agent's identity attributes in the TRUST sense. Just as we track which PKI authority issued a certificate, we must track which model version underlies an agent, who trained or modified it, and what attestation exists for its intended behavior. AI model cards and model provenance records, analogous to the software bill of materials (SBOM) that CISA has championed for traditional software, are the governance mechanism. A model without signed provenance documentation has no established TRUST position.

Supply chain integrity applies to model artifacts. A compromised fine-tune is a supply chain attack. An agent instantiated from a tampered model image is the AI equivalent of running malware with a valid certificate. TRUST for agents means verifying the integrity of the model artifact, the system prompt, and the tool configuration at every instantiation — not just at initial deployment review.

Cross-agent trust must be explicit, scoped, and non-transitive. When an orchestrator spawns a sub-agent, the sub-agent must not automatically inherit the orchestrator's TRUST position. The authority granted must be explicitly declared, narrower than the orchestrator's grant, and time-bounded to the task. Every delegation step must make the TRUST scope smaller, not preserve or expand it.

Synthetic identity is undermining the human anchor at the top of every accountability chain. Before we can govern an agent's TRUST position, we have to be certain the human or organizational sponsor behind it is who they claim to be. That certainty is under direct attack. AI-generated identity documents, synthetic biometrics, and fabricated identity profiles have reached a quality threshold where traditional proofing controls can be defeated at scale, undermining TRUST at the moment of establishment. At runtime, synthetic principals compound the problem: built to appear legitimate from the ground up, they don't exhibit the behavioral inconsistencies that anomaly detection was designed to flag. The architectural response is to move TRUST's anchor from document-based proofing to hardware-backed cryptographic binding, FIDO2/WebAuthn at AAL3, TPM-attested device identity, and to link every runtime principal continuously back to the hardware credential issued at establishment. Synthetic identity cannot replicate a hardware binding it was never issued.

4.2 The TRUST Posture Model: A Compositional Formulation

The TRUST posture model builds on the vectors-of-trust concept established in IETF RFC 8485, which defines component dimensions for describing the trust characteristics of identity transactions between parties,

and extends that lineage from transaction-level signaling to continuous principal-level governance. How these components are evaluated in practice depends on the signals available from deployed products and their integration. No single platform will surface all of them, and organizations should treat this as a governance architecture to work toward rather than a formula to compute precisely.

TRUST posture is composed of four named components, each evaluating a distinct aspect of a principal's identity position.

Human Identity Assurance (HIA) addresses the strength of the human identity claim at the top of every accountability chain. Its key inputs are the PQC migration status of the credential used to establish identity, the quality of liveness detection at proofing time — specifically its resistance to AI-generated synthetic biometrics — and verification that the named sponsor binding remains valid and traceable.

Non-Person Entity Assurance (NPEA) addresses the integrity of non-human principals other than agents — devices, service accounts, and traditional NPEs. Its key inputs are the integrity of the parent anchor co-signature under a PQC-capable algorithm, the integrity of the software bill of materials (SBOM) under a PQC-capable signature, and verified group membership in the appropriate governance groups.

Agent Identity Assurance (AIA) is introduced in this framework as a first-class component, distinct from NPEA, because agents have unique TRUST attributes with no NPE analog. Its key inputs are the integrity of model provenance attestation under a PQC-capable algorithm, the validity of the named sponsor binding, verified agent group membership, and compliance with delegation depth constraints.

System Integrity Assurance (SIA) operates differently from the three components above. It isn't a factor in the TRUST posture assessment; it's a precondition for running the assessment at all. If system integrity falls below an acceptable threshold, as evidenced by secure boot measurement, EDR telemetry, and continuous posture monitoring, TRUST evaluation is suspended entirely, regardless of how strong HIA, NPEA, or AIA may be. A perfectly attested agent running on a compromised host provides no TRUST assurance. System integrity is a gate, not a weight.

Every part of this model has to hold on its own. A strong result in one component does not make up for a weak result in another. A well-issued, hardware-bound credential does not make up for a sponsor nobody has verified. Signed model provenance does not make up for an expired group membership, or a delegation chain that has run deeper than it was authorized to go. These are links in an accountability chain, and a chain does not get stronger because three of its four links are excellent.

That rule, no component covers for another, is part of the framework. The math used to enforce it is not. The two are easy to confuse, and only one of them is settled.

Earlier drafts of this framework assumed you would multiply the three component scores together, with SIA acting as a gate. Multiplying does honor the rule, and it is a reasonable starting point. But it is neither the only way to honor it nor the strictest. Multiplying shrinks a weak score; it does not eliminate it. Three components scoring 0.7, 0.7, and 0.9 multiply out to roughly 0.44, and whether that passes depends entirely on where the bar is set. Two genuinely weak components can survive the calculation.

A simpler approach is often stricter: require each component to clear its own minimum before anything is combined at all. If any one component falls below its floor, the assessment stops there. Weighted averages and other statistical methods can also work, depending on how much each signal is trusted and how independent those signals really are. The right choice depends on what an organization's tools actually measure, and how reliably they measure it.

Organizations should therefore choose the method against their own environment, set their own floors and thresholds rather than inheriting numbers from this paper, and expect to revisit that choice as their signals improve. What should not change is the rule itself, and the SIA gate that comes before it.

SIA is not part of the math. It is a precondition. A perfectly attested agent running on a compromised host tells you nothing about that agent, and no scoring method, however it is weighted, should be able to average that away.

As discussed in Section 7.2, this posture model isn't a static governance artifact. It becomes a live policy instrument that BEHAVE drives continuously, re-evaluating each component as signals arrive from the environment.

4.3 Post-Quantum Requirements for TRUST

MUR and MAR anchoring must migrate to quantum-resistant signatures. All additions, deletions, and modifications to governance records must be anchored using stateful hash-based signatures (XMSS or LMS, per NIST SP 800-208). These provide the long-term signature integrity required for governance records that may be referenced years after creation, with security based on hash function properties that are not threatened by Shor's algorithm.

Model provenance attestations require long-horizon signature integrity. Because model provenance records may be referenced for months or years, they must be signed with ML-DSA (FIPS 204) as the primary algorithm, with governance records requiring archival integrity using stateful hash-based signatures.

System integrity within the SIA gate requires TPM-backed PQC attestation. Firmware measurements and boot attestations signed with LMS or XMSS per NIST SP 800-208 provide the foundation. Continuous posture telemetry must be encrypted with ML-KEM to protect against HNDL collection.

5. CRED: Quantum-Resilient Credentials for Ephemeral, Dynamic Principals

5.1 Classical CRED and Its Evolution for Agents

The CDM CRED capability was built around credentials that are issued, held, and presented: PIV cards, certificates, passwords, derived credentials. NIST SP 800-63B defines the authenticator assurance levels (AAL1–AAL3) and binding requirements that govern how strongly a credential must be tied to its principal before it can be trusted for a given transaction. That's the baseline this paper assumes when it later requires AAL3-equivalent, hardware-backed binding for agent sponsor anchors (Section 4.1) and workload identities. Even as it evolved to cover NPE credentials, the model remained: a credential is issued to a persistent principal, held, and presented to authenticate. For NPEs, the industry generally never evolved beyond shared secrets or static bearer tokens, a governance asymmetry that left non-human identity as the weakest link in most enterprise environments. Agents don't stretch this model; they break it.

An agent may exist for seconds. It may spawn dozens of child processes requiring independent authentication. It may need to authenticate to ten services during a single task. A new instance may be spun up for each invocation with no memory of previous credential exchanges. Static, long-lived credentials in this environment aren't just operationally awkward; they're a security liability. An agent holding a long-lived API key in its context window is a breach surface accessible to the agent's reasoning process, potentially echoed in outputs, potentially extractable via prompt injection.

Workload identity is the correct primitive. The SPIFFE/SPIRE framework issues short-lived, cryptographically attested identity documents (SVIDs) to workloads based on runtime context, not pre-provisioned secrets. An agent's credential should derive from its deployment context: where it is running, what task it was instantiated for, who authorized the instantiation. The IETF WIMSE (Workload Identity in Multi-System Environments) working group, co-chaired by Justin Richer (who also authored RFC 8485), is actively standardizing JOSE-based token solutions for service-to-service HTTP/REST calls and token exchange at security boundaries (building on RFC 8693). WIMSE is the most relevant active standardization effort for agent-to-service authentication in REST-based tool calls, which is the dominant agentic AI interaction pattern today. It doesn't cover every agent credential use case; model provenance attestation and human sponsor anchor governance fall outside its scope. But it is the correct primitive for the tool call authentication layer.

Note that agents authenticating to enterprise services via SAML federation have no standards-defined PQC migration path. The OASIS Security Services (SAML) Technical Committee was closed by Project Administration on 08 July 2023 and is no longer active. There is no standards body to define a SAML PQC algorithm migration, which makes SAML-dependent agent authentication an unresolved risk. This is addressed further in Section 5.2.

OAuth 2.0 with Rich Authorization Requests (RFC 9396) is the access model. When an agent needs to call a tool or API, it should request a scoped access token for that specific operation at the moment it needs it. Combined with short token lifetimes and proof-of-possession binding via DPoP (RFC 9449), this creates a CRED posture that fits agentic workloads. The CDM ICAM Reference Architecture v1.3 identified shared contextual signals as a foundational requirement for continuous evaluation — OIDC's rich claim structure is the protocol-level expression of that requirement, carrying real-time posture signals, step-up authentication context, device compliance state, risk scores, and agent-specific claims that enable the dimensional reassertion model described in Section 7.2.

Delegation chains require hard scope-reduction rules. Credential delegation in multi-agent pipelines must reduce scope at every step. Without explicit depth limits and mandatory scope reduction, you get unbounded privilege propagation — the AI equivalent of a Kerberos golden ticket. Every hop must make the CRED scope smaller.

Delegation chain standards are not yet well-defined. The most developed prior art for policy-governed delegated authorization is UMA 2.0 (User-Managed Access, Kantara Initiative), developed under Eve Maler's leadership, which addressed resource owner delegation to clients with fine-grained policy. UMA 2.0 was designed for human-to-client delegation, not agent-to-agent delegation chains with depth constraints. RFC 8693 (OAuth 2.0 Token Exchange) and GNAP (RFC 9635) provide partial building blocks, but no active IETF working group is currently chartered specifically for multi-agent delegation chain governance with depth limits and mandatory scope reduction. That gap is real, and practitioners should treat delegation chain governance as an area requiring custom policy controls until standards mature.

Agents must never hold human credentials. Giving agents access to human-level credentials — usernames, passwords, session cookies — is an architectural anti-pattern. The correct pattern is OAuth 2.0 delegated authorization with explicit user consent, narrow scope, short-lived tokens, and cryptographic binding to the agent's workload identity.

5.2 Post-Quantum Requirements for CRED

Transport layer hybrid key establishment as the immediate priority. The HNDL threat to credential exchanges is active today. Externally exposed IAM endpoints must adopt hybrid key establishment (ECDH combined with ML-KEM, per FIPS 203) at the TLS layer as the first migration priority. This renders harvested session data mathematically useless to a quantum adversary, even if classical algorithms remain in use at the message layer during the transition period.

Token and assertion signing: ML-DSA as the migration target. JWT signatures on access tokens and ID tokens, and the signing of SPIFFE SVIDs, must migrate to ML-DSA (FIPS 204). The verifier distribution challenge is particularly acute: few signers (IdPs, token services, SPIFFE trust domains) but many distributed verifiers (resource servers, API gateways, relying parties). Migration must proceed as a dual-track strategy — publishing dual-algorithm JWKS endpoints that serve both classical and ML-DSA keys simultaneously, upgrading internal verifiers in waves before deprecating classical algorithms at federation boundaries.

SAML is a forcing function, not a migration target. The OASIS Security Services (SAML) Technical Committee was closed on 08 July 2023. There is no active standards body to define a PQC migration path for XML Signature, and SAML's XML Signature architecture has no clean ML-DSA migration path even if one existed. For federal agencies, M-26-15's Phase 4 signature migration deadline (2031) is the forcing function: every SAML-dependent system must migrate to OIDC/JWT before that deadline, or face an authentication layer that cannot be made quantum-resistant. That's an architectural requirement, not just a cryptographic preference. OIDC is not simply the PQC-compatible alternative to SAML; it is architecturally superior for the identity governance model this paper describes. OIDC tokens carry rich contextual signals (posture data, step-up context, risk scores, agent-specific claims) that the CDM ICAM Reference Architecture identified as foundational for shared signal-based continuous evaluation, and that the dimensional reassertion model in Section 7.2 requires. A SAML assertion tells you who authenticated and when. An OIDC token tells you everything the policy enforcement point needs to make a dimensional reassertion decision rather than a binary deny.

Merkle tree certificates: an emerging efficiency mechanism for high-frequency agent credential verification. Chrome and Cloudflare, in collaboration with the IETF PLANTS working group, are developing Merkle Tree Certificates (MTCs) as an evolution of HTTPS certificate infrastructure. MTCs allow a verifier to confirm a certificate's inclusion in a transparency log with a compact proof, without downloading the entire log. For agentic AI specifically, this matters in a multi-agent pipeline where dozens of short-lived workload identity SVIDs are being issued and verified in rapid succession; efficient certificate status verification becomes an operational requirement. ML-DSA signatures are significantly larger than RSA/ECC equivalents — approximately 3,309 bytes versus 256 bytes for RSA-2048 — compounding bandwidth and processing burden at high verification frequency. Merkle tree-based transparency structures are a promising efficiency mechanism for this use case, and their adoption by Cloudflare and Google signals strong industry direction. Practitioners deploying high-frequency agent credential verification architectures should monitor the PLANTS working group output.

Agent-specific CRED consideration: ephemeral tokens are a quantum advantage. Agentic systems that implement JIT/JEA credential issuance — short-lived, task-scoped tokens discarded at task completion — have a naturally lower HNDL exposure profile. Architectures that are correct from a least-privilege perspective are also better positioned for quantum resilience.

PKI and trust fabric: the longest migration horizon. CA infrastructure must establish a PQC transition roadmap: deploying PQC-capable issuing CAs, re-issuing mTLS certificates with ML-DSA or SLH-DSA signatures, and updating trust stores across the verifier ecosystem. For agentic systems, the SPIFFE trust domain's root CA and the signing keys for SVIDs must be included in the PKI migration plan.

6. PRIV: Governing Privilege That Is Acquired at Runtime

6.1 Classical PRIV and Its Evolution for Agents

Every privilege governance model ever built rests on a single foundational assumption: privilege is determined in advance. You do role engineering, define entitlements, provision access, and certify that what was granted matches what is needed. An agent makes this model structurally inapplicable.

An agent does not have a fixed entitlement set. It has a tool manifest and decides at runtime which tools to call, in what order, with what parameters. Its effective privilege surface at any moment is a function of its task, its context, and its reasoning. You cannot model that as a role. You cannot certify it as a static entitlement.

The tool manifest is the entitlement set, and must be governed as one. Governing agent privilege means governing tool manifests with the same rigor applied to human access grants: who authorizes this manifest, what review is required before adding a tool, what is the blast radius if a tool is misused.

Just-in-time, just-enough-access is the native model. Tool access should be granted at task time, scoped to the specific task, and revoked when the task completes. This is JIT/JEA applied to the AI layer — the same principle CDM PAM implemented for human administrators, now applied to tool authorization at the API or Model Context Protocol (MCP) layer.

Irreversible actions require human authorization in architecture, not just policy. Sending emails to external parties, deleting data, making financial transactions, modifying production infrastructure — these are actions where the blast radius of an error or a prompt injection attack is unacceptable without a human checkpoint.

Separation of duties extends into agent pipelines. No single agent should hold both the authority to approve an action and the capability to execute it. SoD for agents is a pipeline architecture requirement.

Agent access certification is a new governance requirement. Just as human access is periodically certified, agent tool manifests require periodic review: does this agent still require this tool? Has the task scope changed? Has the risk profile of the tool changed? No IGA platform currently provides this capability as a native function.

6.2 Post-Quantum Requirements for PRIV and the Limits of Privilege Specification

PRIV controls for agentic AI are necessary but provably insufficient, and that limitation is mathematical, not operational. NIST researcher Apostol Vassilev has demonstrated, applying Gödel's incompleteness theorems, that no finite set of guardrails can be universally robust against adversarial prompts (Robust AI Security and Alignment: A Sisyphean Endeavor, IEEE Security & Privacy, May 2026). The same logic applies directly to privilege specification. No finite tool manifest or authorization policy can fully enumerate the intent boundary of a non-deterministic reasoner operating in an open environment. There will always be a task context, a tool invocation sequence, or an adversarial prompt that the privilege specification did not anticipate. That is the irreducible residual risk of agentic AI governance, and no governance framework can eliminate it. The correct architectural response is to acknowledge that limit explicitly and put the compensating work where it belongs: on BEHAVE. That is a heavier burden than BEHAVE has had to carry for any prior principal type, and the framework needs to be honest that it now depends on BEHAVE holding up its end.

Ephemeral, time-decayed privilege tokens are the correct PQC-aligned pattern. JIT provisioning paired with automated privilege decay drastically reduces the HNDL value of any individual authorization assertion. A privilege token that expires in minutes is a low-value harvest target.

Policy Decision Point integrity requires PQC-protected policy stores. Policy bundles — OPA policy bundles, Cedar policy sets — must migrate to PQC-signed artifacts to prevent quantum-era tampering.

Algorithm confusion attacks must be addressed at privilege enforcement points. During the migration period, when both classical and post-quantum algorithms are accepted, PEPs must implement strict algorithm binding. A token that claims ML-DSA signing must fail validation if its signature validates only under RS256.

7. BEHAVE: Runtime Conformance in the Presence of Cryptographic Transition

7.1 Classical BEHAVE and Its Evolution for Agents

NIST researcher Apostol Vassilev has provided the mathematical foundation for why BEHAVE matters so much for agentic AI governance: it isn't merely useful, it's a provable necessity. His proof, published in IEEE Security & Privacy (May 2026, DOI 10.1109/MSEC.2026.3678214), demonstrates using Gödel's incompleteness theorems that no finite set of guardrails can be universally robust against adversarial prompts. Just as no finite set of axioms can produce a complete and consistent mathematical system, no finite set of behavioral constraints on an AI system can be complete and consistent against all possible adversarial inputs. Section 6.2 established that this makes PRIV provably insufficient for agents. The Vassilev proof makes the implication for BEHAVE equally precise. Continuous monitoring and updating isn't a best practice for agentic AI governance; it's a mathematical requirement.

Two implications follow directly from the proof.

First, BEHAVE must be a continuous monitor-and-update capability, not a static anomaly detection model. A fixed behavioral baseline for an agent is subject to the same incompleteness as a fixed guardrail set. Adversaries will find the prompt that defeats it. The correct model is continuous red teaming of agent behavioral policies combined with continuous updates to detection logic.

Second, the BEHAVE objective for agents is economic, not absolute. Vassilev is explicit: complete security is impossible, but you can make exploitation financially prohibitive. BEHAVE must be calibrated to raise the cost of successful behavioral manipulation above the attacker's resource threshold, not to achieve perfect detection, which the proof shows is unattainable.

The classical BEHAVE model must also be re-anchored for agents. UEBA was designed around behavioral baselines — historical patterns of access, time-of-day norms, peer group comparisons. An agent has no meaningful historical baseline: it is designed to behave differently based on its input. The correct anchor is declared purpose, not historical pattern. For humans, BEHAVE asks: Is this principal acting like themselves? For agents: Is this agent acting within the scope of what it was instantiated to do?

Tool call sequences are the primary behavioral signal. The sequence, parameters, and targets of tool calls are to agents what keystrokes and file access patterns are to humans. An agent querying internal HR systems

and enumerating user accounts in the context of a document summarization task is behaving anomalously — not because it lacks technical permission, but because it is acting outside its declared purpose.

Prompt injection is a BEHAVE signal, not just an input validation problem. When an agent begins taking actions inconsistent with its task context following ingestion of external content, this is a behavioral anomaly requiring BEHAVE controls: session suspension, escalation to the named sponsor, and forensic capture of the agent's context.

Audit trails must be continuous, complete, and structured. BEHAVE requires evidence. For agents this means logging every tool call, every external invocation, every sub-agent spawned, and where feasible the reasoning context that preceded consequential actions. OMB Memorandum M-26-14, the companion logging and visibility mandate issued alongside M-26-15's PQC migration requirements, is directly on point here. It establishes the federal baseline for what must be logged, retained, and made available for continuous evaluation, and agent tool-call telemetry should be treated as falling squarely within its scope rather than as a novel logging category exempt from existing policy.

Behavioral correlation must operate at the pipeline level. A single agent session may look innocuous. A pattern of sessions — same orchestrator, progressively accessing more sensitive resources — may constitute an attack in progress. BEHAVE for agents is a pipeline-level, cross-session analytical requirement.

7.2 Dimensional Reassertion: BEHAVE as the Continuous Verification Signal

The four dimensions are not independent monitoring silos. They form a continuous feedback loop, and that isn't merely an architectural preference. It is a requirement of Zero Trust Architecture and the context-based access control model that has been central to CISA publications, including the CDM ICAM Reference Architecture v1.3. NIST SP 800-207 is explicit: access decisions must be continuously re-evaluated based on current context, not just initial authentication state. The policy enforcement point doesn't simply allow or deny; it evaluates the full context of the request against current policy and current principal posture.

When BEHAVE detects an anomaly, the ZTA continuous verification requirement mandates dimensional reassertion, not reflexive binary denial. Denial without reassertion wastes the most valuable thing BEHAVE just produced: a precise signal about which dimension drifted and by how much. Dimensional reassertion uses that signal. Binary denial discards it.

The correct response model has three tiers, calibrated to signal strength.

Drift — non-deterministic behavior expressing a non-preferred path within the agent's declared purpose. Non-deterministic systems drift. This is not necessarily an indication of compromise; it may simply represent the irreducible non-determinism that Vassilev's proof establishes as mathematically unavoidable. The appropriate response is minimum necessary intervention: session-level reassertion, the agent's context reset, a new session instantiated with a clean state. The principal's TRUST position is maintained. The named sponsor is notified as an informational event. This mirrors what an experienced practitioner does in practice: stop the chat, not stop using the system.

Deviation — behavior outside the agent's declared purpose but without confirmed compromise indicators. The appropriate response is session termination and restart, MAR flagged for governance review, named sponsor notified for acknowledgment, tool manifest scope temporarily reduced pending review. No credential revocation. TRUST position maintained pending review outcome. The TRUST posture is re-evaluated with the AIA component reduced pending sponsor confirmation.

Compromise — active exploitation indicators, confirmed prompt injection, credential misuse, or actions that cannot be explained by declared purpose under any reasonable interpretation. The appropriate response is hard deny, immediate credential revocation, MAR suspended, sub-agents orphaned and terminated, named sponsor escalated for incident response. The TRUST posture is zeroed until full re-establishment under sponsor authorization.

This tiered model maps directly to the TRUST posture model introduced in Section 4.2. When BEHAVE triggers reassertion, it is driving a live re-evaluation of that posture. A Deviation event reduces the AIA component pending sponsor review. A Compromise event zeroes the AIA component entirely, and under

any composition that satisfies the no-compensation rule, zeroing one component zeroes the aggregate. So BEHAVE does more than watch for problems after the fact: it is the mechanism that keeps the other three dimensions calibrated to current reality, continuously. This also fits the economic objective Vassilev identifies. Targeted reassertion raises the cost of continued exploitation without revealing exactly what was detected, which is a more resilient posture than binary denial, since denial hands the adversary a clean confirmation signal.

7.3 Post-Quantum Requirements for BEHAVE

BEHAVE as a cryptographic transition safety net. The transition to post-quantum cryptography introduces significant implementation complexity, which historically correlates with exploitable vulnerabilities. Misconfigurations, algorithm confusion vulnerabilities, and implementation flaws in PQC libraries will be exploited before they are discovered. BEHAVE is the enterprise's primary detection capability when cryptographic controls fail: anomalous access velocity, unexpected API invocation patterns, and unusual data extraction rates may all indicate that a compromised or poorly implemented PQC credential is being exploited. For agents, behavioral anomaly detection must be calibrated against the declared task scope rather than historical behavior — the expected tool call rate and invocation pattern for a document summarization task looks nothing like that for a data pipeline agent, and thresholds must reflect those task-specific norms rather than population averages. How that calibration is implemented will depend entirely on the monitoring and analytics tooling available in a given environment; the governance requirement is that the calibration exist and be reviewed when the agent's declared purpose changes.

Audit log integrity requires quantum-resistant signing. Agent audit logs and all MAR governance records must be signed with stateful hash-based signatures (LMS or XMSS) that provide long-horizon integrity guarantees. This is the same requirement as MUR anchoring in the TRUST dimension: governance records have long lifetimes and require correspondingly long-lived signature integrity.

BEHAVE monitoring must account for PQC transition-specific attack patterns. During the migration period, BEHAVE should be configured to detect unusual cryptographic algorithm negotiation patterns (possible downgrade attacks); access events using legacy algorithms from principals whose records indicate they should have migrated; and sudden changes in authentication latency patterns that might indicate injection of a PQC-incompatible client into a migrated flow.

8. The Master Agent Record: Operationalizing the Four Dimensions

8.1 Concept and Purpose

The CDM framework's most powerful operational artifact for human identity is the Master User Record (MUR): a unified representation of a principal's identity attributes, credential status, privilege grants, and behavioral posture. Agentic AI governance requires the direct equivalent: a Master Agent Record (MAR).

The MAR is a governed identity artifact maintained in an authoritative identity store, subject to lifecycle management, continuously updated with operational telemetry, and reviewed by a named sponsor on a defined cadence. It operationalizes all four dimensions for an agent principal, including PQC governance state across each dimension.

8.2 MAR Attribute Schema

Identity and accountability attributes (TRUST dimension): Unique agent identifier (immutable, issued at instantiation); named sponsor name or organizational role with traceable accountability chain (must resolve to a human or organizational anchor — another agent is not a valid sponsor); agent group memberships — the governance groups this agent belongs to, driving policy inheritance, access certification scope, and SoD enforcement (group membership may be task-context-sensitive — an agent instantiated for a high-sensitivity task may temporarily inherit a more privileged group context, which is an emerging governance pattern without direct human analog); organizational owner and deployment authority; deployment date and intended purpose statement (specific enough for behavioral baseline derivation); instantiation authorization record.

Model provenance attributes (TRUST dimension — PQC-protected): Foundation model name, version, and provider; fine-tune lineage if applicable; system prompt hash (signed with ML-DSA at deployment); tool manifest hash (signed with ML-DSA at deployment); model provenance attestation (signed with LMS/XMSS for long-horizon integrity); last provenance verification timestamp and verifying authority.

Credential profile (CRED dimension — PQC posture tracked): Workload identity type (SPIFFE SVID or equivalent); credential issuing authority and trust domain; current PQC algorithm posture (classical-only/hybrid/PQC-native); token scope manifest; credential rotation policy and last rotation timestamp; delegation depth limit and current delegation chain record.

Privilege profile (PRIV dimension): Authorized tool manifest with parameter constraints; maximum privilege scope and least-privilege justification; actions requiring human-in-the-loop authorization; blast radius assessment for each authorized tool; last access certification date and certifying reviewer; SoD constraints applicable to this agent's role in multi-agent pipelines.

Behavioral policy (BEHAVE dimension — PQC audit integrity tracked): Declared task scope in plain language (the behavioral anchor); behavioral monitoring policy reference; anomaly tier thresholds (Drift/Deviation/Compromise) and escalation paths; named sponsor notification thresholds and contact; audit log location, retention policy, and signing algorithm; PQC migration status of audit log signing.

PQC governance state (cross-cutting — Vikram's MV-QRIAM requirements status): Current cryptographic posture across transport, token, and trust fabric layers; Vikram's MV-QRIAM requirements status (R0–R5 assessment); migration phase and target completion dates aligned to M-26-15 federal timeline; outstanding PQC remediation items with priority classification.

8.3 MAR Lifecycle

Instantiation — Named sponsor authorization obtained; MAR created and signed with ML-DSA; model provenance attested and anchored with LMS/XMSS; tool manifest approved by governance reviewer; agent group memberships assigned; workload identity credential issued; initial PQC posture assessment recorded. No agent reaches production without a complete, signed MAR.

Modification — Tool manifest changes, scope expansions, model version updates, group membership changes, and PQC migration milestones each trigger governance review proportional to risk. Every modification record is signed and linked to the authorization decision.

Suspension — Anomalous behavior (Compromise tier), named sponsor departure, or security incident triggers immediate credential revocation and tool access suspension. The MAR is placed in suspended state pending review.

Decommission — Credentials revoked, tool access removed, MAR archived with full revision history, audit log preserved for the compliance-required retention period.

9. Migration Playbook: Harmonizing M-26-15, Vikram's MV-QRIAM, and Agent Identity Governance

The migration playbook that follows aligns to OMB M-26-15's five-phase federal timeline as the governing framework, with Vikram's MV-QRIAM technical sequencing logic preserved within that structure. Federal agencies are bound by M-26-15's hard dates; non-federal organizations should treat the sequence as a reference architecture rather than a compliance deadline. Within each phase, actions are organized across three tracks:

- **All Principals:** IAM modernization actions that serve human and agent identity simultaneously
- **Agent-Specific Extensions:** additions to existing IAM motions that extend coverage to agent principals
- **Net-New Requirements:** governance capabilities with no human identity analog

ICAM modernization requires that both the PQC migration workstream and the agent identity governance workstream be included in the same program plan. Executing them independently will force a second pass

through the same infrastructure and miss the efficiency opportunities that the simultaneous action structure below makes explicit.

One note on the ICAM signature gap: CISA has formally classified ICAM as transitioning. Key establishment (ML-KEM) is progressing, but digital signature migration (ML-DSA) is lagging significantly. Organizations must not mistake completion of Phase 2 transport modernization for ICAM PQC readiness. The authentication layer, every token signature, every federation assertion, every certificate, remains classically vulnerable until Phase 4 signature migration is complete.

Phase 1 — Strategy, Planning, and Discovery (2026–2027)

All Principals: Execute automated cryptographic inventory (CBOM) of all IAM infrastructure per M-26-15 automation requirements — PKI, token issuance, directory services, federation endpoints. Human and agent identity share this infrastructure; one inventory serves both workstreams. Conduct principal inventory: extend existing MUR audit processes to enumerate all agent deployments — same governance motion, different principal type. Assess IGA platform readiness for both PQC-capable credential management and non-human principal governance simultaneously. Engage GSA FICAM inter-agency working group (icam@gsa.gov): the M-26-15 mandated coordination mechanism for FICAM modernization, relevant to both human identity PQC migration and agent identity governance standards. Designate PQC Migration Program Manager and Security Architect per M-26-15 Appendix B roles. Submit PQC Migration Plan to OMB/ONCD within 120 days of June 24, 2026 (deadline approximately October 22, 2026).

Agent-Specific Extensions: Assign named sponsors for all inventoried agents, the accountability anchor that makes agent inventory governable. Create proto-MAR from inventory data: a structured registry establishing governance discipline before tooling exists. Define an agent group membership schema. What groups will agents belong to, what policies will that membership drive, and how will task-context-sensitive membership be governed?

Net-New Requirements: Tool manifest audit and least-privilege assessment for all deployed agents. Declared purpose documentation for each agent (the BEHAVE behavioral anchor). Vikram's MV-QRIAM R0 cryptographic assessment scoped to agent-specific credential surfaces.

Phase 2 — Pilots and Early Migration (2027–2028)

All Principals: Deploy hybrid TLS 1.3 (X25519 + ML-KEM-768) at all IAM endpoints — SSO flows, federation, and agent tool call traffic share the same gateway infrastructure; one modernization serves all principals. Engage GSA FICAM pilot program for PQC-ready physical and logical access systems. Begin IGA platform upgrades to support non-human principal governance alongside PQC-capable credential management — these are the same platform modernization effort. Publish dual-algorithm JWKS endpoints for IdPs (RS256 + ML-DSA-65): serves both human and agent token verification.

Agent-Specific Extensions: Deploy WIMSE-based workload identity for agents, aligned with token issuance infrastructure already under upgrade for human identity. Implement JIT/JEA tool authorization framework — the agent-native privilege model. Formalize MAR schema and begin systematic MAR creation for all production agents. Begin agent access certification capability development in IGA platform.

Net-New Requirements: Cross-agent delegation chain depth limit enforcement. Human-in-the-loop checkpoint architecture for irreversible agent actions. Agent behavioral monitoring baseline deployment using declared purpose framework. Vikram's MV-QRIAM R1–R2 compliance assessment.

Phase 3 — Prioritized Migration for Key Establishment (2028–2030)

Deadline January 2, 2030 — TLS 1.3 and ML-KEM key establishment required for all High Value Assets per M-26-15.

All Principals: Complete ML-KEM key establishment migration for all HVAs — IAM infrastructure is explicitly in scope. Verify cryptographic agility across all IAM components. Upgrade internal verifier ecosystem to accept ML-DSA-signed tokens (addressing Vikram's MV-QRIAM signer-verifier asymmetry — few signers, many distributed verifiers). Migrate SCIM provisioning payloads to PQC-protected transport.

Agent-Specific Extensions: Migrate SPIFFE trust domain root CA to PQC-capable signing. Re-issue agent SVIDs under PQC-signed trust domain. Deploy fully-formed MAR schema with ML-DSA-signed model provenance attestation. Launch agent access certification campaigns in IGA platform. Evaluate Merkle tree certificate infrastructure for high-frequency agent credential verification — monitor IETF PLANTS working group output.

Net-New Requirements: Cross-agent behavioral correlation capability deployment. Pipeline-level BEHAVE monitoring across multi-agent workflows. Vikram's MV-QRIAM R3–R4 compliance assessment for agent identity plane.

Phase 4 — Signature Migration (2031)

Deadline 2031 — ML-DSA signature migration required for all HVAs per M-26-15.

This phase is the SAML forcing function. The OASIS SAML Technical Committee closed on 08 July 2023. There is no active standards body to define a PQC migration path for XML Signature, and SAML's architecture has no clean ML-DSA migration path. Every SAML-dependent system (human SSO, agent federation, any remaining SAML assertion flow) must complete migration to OIDC/JWT before this deadline, or face an authentication layer that cannot be made quantum-resistant. OIDC is not simply a SAML replacement. It provides the rich contextual signals, posture data, step-up context, risk scores, agent-specific claims, that the continuous evaluation and dimensional reassertion model requires. SAML tells you who authenticated. OIDC tells you everything the policy enforcement point needs to make a dimensional reassertion decision.

All Principals: Complete ML-DSA signature migration for all HVAs — JWT signing, all remaining SAML assertion flows replaced with OIDC, certificate signing. Complete PKI root CA re-establishment under PQC algorithms. Migrate MUR anchoring to LMS/XMSS stateful hash-based signatures for long-horizon governance record integrity.

Agent-Specific Extensions: Migrate MAR governance record signing to LMS/XMSS. Re-sign model provenance attestations under archival-grade PQC algorithms. Migrate agent audit log chain signing to PQC.

Net-New Requirements: Full Vikram's MV-QRIAM compliance certification (R0–R5) for agent identity plane. Re-validate TRUST posture composition under fully PQC-native infrastructure. BEHAVE monitoring updated to detect post-migration PQC-specific attack patterns.

Phase 5 — Full Migration and Continuous Posture (2035)

Deadline 2035 — Full PQC migration required per M-26-15.

All Principals: Complete migration of remaining systems based on risk assessment and commercial availability. Deprecate classical algorithms at all IAM enforcement points. Implement continuous cryptographic posture monitoring via ACDI tools per M-26-15.

Agent-Specific Extensions: MAR as a standard governed artifact across all agency agent deployments. IGA platforms natively supporting agent access certification, group governance, and MAR lifecycle management. Agent behavioral governance integrated into CDM dashboard.

Net-New Requirements: Dimensional reassertion capabilities (Drift/Deviation/Compromise tiers) fully operationalized. TRUST posture model operating as a live policy instrument driving the continuous BEHAVE feedback loop. Agent identity governance framework published as agency standard and contributed to GSA FICAM working group for inter-agency adoption.

10. What Organizations Should Do Now

Waiting for standards to mature is not an option. Agents are in production. HNDL collection is active. Governance frameworks are absent. Here is a minimum viable posture for organizations deploying agents today, before the playbook above is underway.

Inventory every agent and assign named sponsors. You cannot govern what you cannot see. A registry of every agentic system in use — including tools, task scope, and named sponsor — is your proto-MAR. No named sponsor means no production deployment. This costs nothing and can be implemented today.

Define agent group memberships. Even before IGA platforms support agent governance natively, define which groups your agents belong to and what governance implications those memberships carry. Group membership is how you scale access certification, SoD enforcement, and policy inheritance to agent populations.

Audit tool manifests for least-privilege compliance. For every deployed agent, document what tools it can invoke, what those tools can do, and the blast radius of misuse. Remove any tool access that is not demonstrably necessary.

Enforce human-in-the-loop checkpoints for irreversible actions in architecture, not policy. If your agents can send external communications, delete data, or modify infrastructure without human confirmation, you have an unacceptable blast radius. Change the architecture.

Start logging agent tool calls with structure. Every tool call, every external invocation, every sub-agent spawned. If you cannot reconstruct what an agent did, you cannot investigate incidents. Begin with structured logging before PQC audit log signing is in place — the content is more urgent than the signature in the short term.

Conduct a PQC inventory for IAM infrastructure, including agent credential surfaces. Catalog which algorithms currently sign your identity tokens, federation assertions, workload identity credentials, and audit records. This is Vikram's MV-QRIAM R0, and you cannot plan a migration without it. Include agent-specific credential surfaces in scope; they are not covered by most existing PQC inventory approaches.

Engage the GSA FICAM inter-agency working group. M-26-15 directs GSA to establish this working group to modernize FICAM for PQC. It is the appropriate institutional home for both the human identity PQC migration and the agent identity governance standards that do not yet exist. Contact icam@gsa.gov.

Apply all four dimensions to every new agent deployment. Before production: Can you characterize this agent's TRUST position, including model provenance? Its CRED posture and PQC migration status? Its PRIV scope and tool manifest? Its declared BEHAVE purpose, anomaly tier thresholds, and audit logging configuration? If you cannot answer all four across both the governance and cryptographic dimensions, the agent is not ready.

11. Conclusion: The Framework Holds, The Work Is Urgent

I spent years arguing that identity is not a feature of enterprise security. It is the control plane. The four dimensions of TRUST, CRED, PRIV, and BEHAVE were built to reflect that. They were designed to be comprehensive: every identity-relevant question about every principal in the enterprise maps to one of these four dimensions.

Post-quantum risk and agentic AI stress-test that claim simultaneously. But they do not break it.

Every cryptographic vulnerability that quantum computing exposes maps to one of the four dimensions. Every governance challenge that agents introduce maps to one of the four dimensions. The sponsorship, provenance, and synthetic identity questions are TRUST problems. The ephemeral credential, delegation chain, and PQC migration questions are CRED problems. The dynamic tool authorization, blast radius, and privilege decay questions are PRIV problems. The behavioral anomaly detection, declared purpose governance, dimensional reassertion, and audit log integrity questions are BEHAVE problems.

What the field lacks is not a new framework. What it lacks is the tooling, the standards, and the institutional discipline to apply the existing framework to new principal types and a changing cryptographic landscape simultaneously. The Master Agent Record does not exist as a standard artifact. Workload identity for AI agents is not in most deployment playbooks. Tool manifest governance is not a feature in any IGA platform. Agent-aware behavioral analytics is at the research stage. PQC migration for agent-specific credential surfaces is not addressed in existing migration guidance. And the SAML-to-OIDC migration that the 2031 signature deadline forces is the single most consequential near-term ICAM modernization action. Not because OIDC is simply more secure, but because OIDC provides the rich contextual signals that continuous evaluation and dimensional reassertion require.

The mathematical proof is in. The policy mandates are issued. The framework is ready. The work is ahead.

The framework holds. The urgency is real.

The author, Ross Foard, is a former senior identity architect with DHS/CISA, where he contributed to the Continuous Diagnostics and Mitigation (CDM) program's Identity and Access Management capability, including original development of the TRUST/CRED/PRIV/BEHAVE framework presented in November 2016 and co-authorship of the CDM ICAM Reference Architecture v1.3 (September 2023). He is President of Foard Consulting LLC and co-author, with Kelvin Brewer of Ping Identity, of "Defensible Zero Trust: A Strategic Playbook" (ATARC, April 2026). Updated on August 26, 2026.

References

- Office of Management and Budget, *Execution of the Migration to Post-Quantum Cryptography*, M-26-15, June 2026
- Office of Management and Budget, M-26-14, June 2026 (federal logging and visibility requirements)
- Executive Order 14412, *Securing the Nation Against Advanced Cryptographic Attacks*, June 2026
- CISA, *Product Categories for Technologies That Use Post-Quantum Cryptography Standards*, January 2026 (pursuant to EO 14306, June 2025)
- CISA, *CDM ICAM Reference Architecture*, Version 1.3, September 2023
- CISA, *Zero Trust Maturity Model*, Version 2.0, April 2023
- NIST, *Module-Lattice-Based Key-Encapsulation Mechanism Standard* (ML-KEM), FIPS 203, August 2024
- NIST, *Module-Lattice-Based Digital Signature Standard* (ML-DSA), FIPS 204, August 2024
- NIST, *Stateless Hash-Based Digital Signature Standard* (SLH-DSA), FIPS 205, August 2024
- NIST, *Stateful Hash-Based Signature Schemes*, SP 800-208, October 2020
- NIST, *Transition to Post-Quantum Cryptography Standards*, IR 8547, Initial Public Draft, November 2024
- NIST, *Considerations for Achieving Cryptographic Agility*, CSWP 39, 2025
- NIST NCCoE, *Mappings of Migration to PQC Project Capabilities to Risk Framework Documents*, CSWP 48, Initial Public Draft, September 2025
- NSA, *Commercial National Security Algorithm Suite 2.0*, September 2022
- Vassilev, A., *Robust AI Security and Alignment: A Sisyphean Endeavor?*, IEEE Security & Privacy, May 2026, DOI 10.1109/MSEC.2026.3678214
- NIST, *Mathematical Proof Supports Transition to a Continuous-Monitor-and-Update Security Model for AI Systems*, News Release, June 2026
- Kalekar, V., *Quantum-Resilient Identity and Access Management (IAM): A Practical Integration Blueprint*, IJSRCSEIT, Volume 12, Issue 3, May–June 2026
- NIST SP 800-207, *Zero Trust Architecture*, August 2020
- NIST SP 800-63B, *Digital Identity Guidelines: Authentication and Lifecycle Management*
- NIST SP 800-63A, *Digital Identity Guidelines: Identity Proofing and Enrollment*
- Foard, R. and Brewer, K., *Defensible Zero Trust: A Strategic Playbook*, ATARC Member Contribution, April 2026
- GSA, *Digital Worker Identity Playbook*, Version 1.1
- NIST AI RMF 1.0, *Artificial Intelligence Risk Management Framework*, January 2023
- IETF RFC 9396, *OAuth 2.0 Rich Authorization Requests*
- IETF RFC 9449, *OAuth 2.0 Demonstrating Proof of Possession (DPoP)*
- IETF RFC 9635, *Grant Negotiation and Authorization Protocol (GNAP)*
- IETF RFC 8705, *OAuth 2.0 Mutual-TLS Client Authentication and Certificate-Bound Access Tokens*
- IETF RFC 8693, *OAuth 2.0 Token Exchange*
- IETF RFC 8485, *Vectors of Trust*, January 2019
- IETF RFC 7519, *JSON Web Token (JWT)*
- IETF RFC 8446, *The Transport Layer Security (TLS) Protocol Version 1.3*
- IETF WIMSE Working Group, *Workload Identity in Multi-System Environments*, datatracker.ietf.org/wg/wimse/about

- IETF PLANTS Working Group, *Merkle Tree Certificates for TLS*, datatracker.ietf.org
- OASIS Security Services (SAML) Technical Committee, closed by Project Administration 08 July 2023, oasis-open.org/committees/security
- Kantara Initiative, *User-Managed Access (UMA) 2.0*, 2017
- SPIFFE Project, *Secure Production Identity Framework for Everyone*, spiffe.io
- NIST SP 800-204D, *Strategies for the Integration of Software Supply Chain Security in DevSecOps*