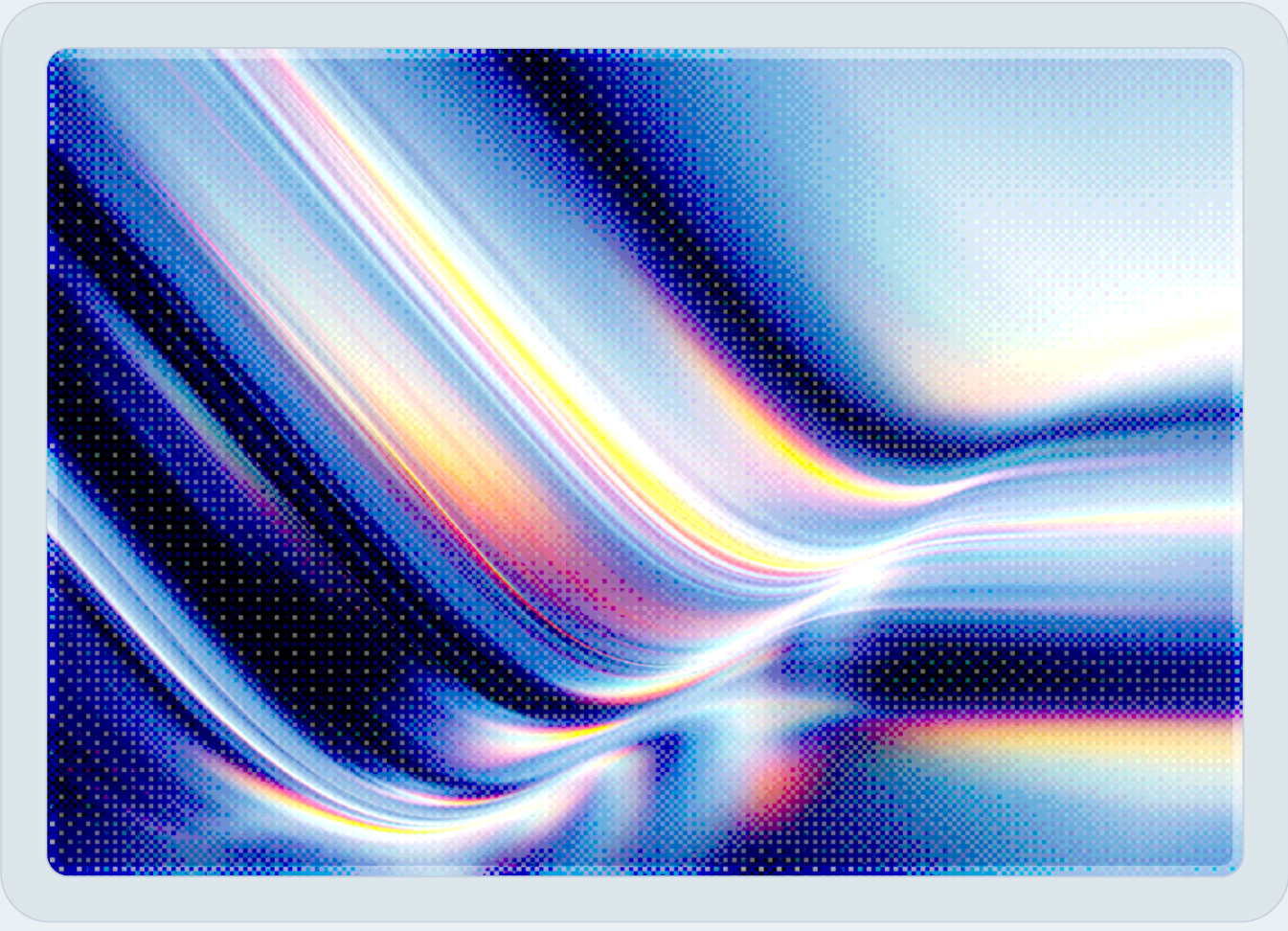
CSAI Foundation | Cloud Security Alliance

NIST IR 8587 and the AI Agent Token Security Gap

Federal Token Protection Guidance Meets a Non-Human Identity Boom

2026-09-26

AI-assisted Rapid Research



CSAI

CSA cloud security alliance®

Key Takeaways

NIST and CISA finalized NIST Interagency Report 8587, "Protecting Tokens and Assertions from Forgery, Theft, and Misuse," on September 15, 2026, giving federal agencies and cloud service providers concrete implementation guidance for defending the identity tokens, access tokens, and assertions that underpin modern single sign-on and API access [1][2][3]. The report responds directly to a documented incident in which foreign state-sponsored actors used a stolen cryptographic signing key to forge authentication tokens and read more than 60,000 emails from a federal agency [1][2], an episode widely understood to reference the 2023 Storm-0558 compromise of Microsoft's consumer signing key and its downstream effect on Exchange Online accounts at the U.S. State Department and other agencies [4]. NIST IR 8587 extends its scope beyond the human-workforce scenarios that motivated it, explicitly addressing token exposure in CI/CD pipelines, machine-to-machine communication, and agentic AI use cases where autonomous agents present signed tokens to reach tools, data, and external services [5]. That extension coincides with a period in which non-human identities, including AI agents, have proliferated faster than the governance practices needed to manage them: a January 2026 survey of 383 IT and security professionals conducted by the Cloud Security Alliance and Oasis Security found that 78 percent of organizations have no documented, formally adopted policy for creating or removing AI identities, and separate industry commentary has flagged legacy identity and access management systems as poorly equipped to handle the speed and volume at which agentic systems generate new machine identities [6] [7]. The gap this note examines is that NIST IR 8587 gives agencies and cloud providers a rigorous playbook for protecting the tokens themselves, while the governance practices needed to know which tokens exist, who or what holds them, and how long they should live remain immature across the non-human identity population the guidance is now explicitly trying to cover.

Background

Access tokens and assertions are the connective tissue of modern cloud authentication and authorization. A user or service authenticates once, typically to an identity provider, and receives a signed token that downstream applications, application programming interfaces, and federated services trust without re-verifying the original credential. This design is efficient, but it concentrates enormous value in two places: the cryptographic keys that sign tokens, and the token verification logic that downstream services trust. NIST IR 8587 was drafted in direct response to a demonstration of what

happens when both of those trust anchors fail simultaneously. In the incident NIST references, a threat actor obtained a Microsoft cryptographic signing key and exploited a flaw in token validation logic to forge authentication tokens that granted access well beyond what the compromised key was originally intended to authorize, ultimately enabling the exfiltration of more than 60,000 emails from a U.S. federal agency [1][2]. That episode, publicly attributed to the China-linked actor tracked as Storm-0558, exposed how a single compromised signing key and an unnoticed validation gap can cascade into a broad, hard-to-detect intrusion across an entire federated identity ecosystem, since forged tokens are, by design, indistinguishable from legitimate ones once they pass verification [4].

NIST released an initial draft of the guidance in December 2025 and finalized it on September 15, 2026, following a public comment period, with Ryan Galluzzo of NIST's Digital Identity Program and Andrew Regenscheid of NIST co-authoring alongside Stephanie Nelson of Accenture Federal Services and Christine Lazcano of CISA [1][3]. The document is formally titled "Protecting Tokens and Assertions from Forgery, Theft, and Misuse: Implementation Recommendations for Agencies and Cloud Service Providers," and it builds on the identity and access control updates made in NIST Special Publication 800-53, Release 5.1.1, as well as the mandates of Executive Order 14306 on federal cybersecurity practices [1][3]. Although the guidance is written primarily for federal agencies and the cloud service providers that serve them, Galluzzo has publicly emphasized that the recommendations are intended for any organization that issues, verifies, or relies on identity tokens, extending its practical relevance well beyond the federal sector [1].

Structurally, NIST IR 8587 organizes its recommendations around the token lifecycle rather than around any single product or protocol. It covers key management practices for the cryptographic material used to sign tokens, architectural guidance for identity providers and authorization servers, protections specific to single sign-on and federation scenarios, controls for application programming interface access, and continuous monitoring practices intended to catch forged or misused tokens before they cause the kind of prolonged, undetected access seen in the 2023 incident [3][5]. Two forward-looking additions distinguish this report from earlier federal token guidance. First, it introduces explicit considerations for AI agent identity, acknowledging that autonomous agents increasingly authenticate to tools, data sources, and third-party services using the same signed-token mechanisms built for human users and traditional service accounts [5]. Second, it directs agencies and cloud providers to begin inventorying their public-key cryptography now and to prepare for post-quantum migration, warning that the larger key and signature sizes associated with algorithms such as ML-KEM and ML-DSA may strain the size assumptions built into JSON Web Tokens, browser cookies, and HTTP headers, and that hardware security modules and key management services configured for classical cryptography will need re-evaluation before they can handle post-quantum operations efficiently [5].

Security Analysis

NIST IR 8587's central contribution, in CSA's reading, is that it treats token forgery and misuse as an architectural problem rather than a checklist item, and its recommendations read as a direct response to the specific failure modes exposed by the Storm-0558 incident. Where that incident succeeded because a single signing key carried far more implicit trust than its issuers intended, and because token validation logic accepted tokens it should have rejected, the guidance pushes agencies toward outcome-focused, less prescriptive key protection requirements, stronger verification procedures, and tighter scoping of what any given key or token is permitted to authorize [1][3]. In CSA's assessment, this is an appropriate response to the specific incident that motivated it, and it reflects a broader industry shift toward treating signing keys as high-value, narrowly scoped secrets rather than durable infrastructure that can be trusted indefinitely once issued.

Where the guidance's ambition runs ahead of current practice is in its extension to AI agent identity. NIST IR 8587 acknowledges that agents present signed tokens to reach tools, data, and services, and it recommends applying the same forgery, theft, and misuse protections to those tokens that apply to human and traditional service-account tokens [5]. That is a reasonable requirement to state, but the population of non-human identities the requirement now covers is large, fast-growing, and, by the account of recent industry surveys, poorly governed. The CSA/Oasis Security survey cited above found that 78 percent of organizations have no documented policy for creating or removing AI agent identities [7], and separate industry commentary has identified authentication and authorization patterns in Model Context Protocol deployments as a specific point of exposure, warning that granular OAuth consent flows and constrained delegation still leave room for token misuse even when applied as designed [6]. Those findings describe a governance environment where organizations frequently cannot answer the basic questions NIST IR 8587's monitoring and lifecycle recommendations presuppose they can answer: which tokens exist, what created them, what they are scoped to do, and when they should be revoked. A key protection standard applied against an inventory that does not exist is necessary but not sufficient; the token could be perfectly protected cryptographically and still represent an unmanaged risk if no one is tracking that it exists or verifying that its scope still matches its original purpose.

A related failure mode specific to agentic systems is the practice, documented in CSA's own research on shadow AI, of relying on a single, broadly scoped API key as the primary authentication mechanism across every agent instance an organization deploys, so that the compromise of one agent effectively compromises the credential set of all of them. One 2026 industry survey cited in that research found that only 22 percent of teams treat agents as independent identities, with most organizations instead depending on shared API keys that provide no granular auditability [8]. By extension, this pattern likely undermines the scoping and lifecycle discipline NIST IR 8587 calls for, since a credential shared across dozens of agent instances cannot meaningfully be tied to a single accountable identity, cannot be

revoked without disrupting every agent that depends on it, and often persists far longer than the short credential lifetimes the guidance recommends because rotating it requires coordinated updates across every consuming agent. Agentic systems compound this risk further because they can dynamically acquire permissions at runtime, spawn subordinate agents, and invoke external application programming interfaces on their own initiative, which means the blast radius of a single forged or stolen agent token can expand in ways that are difficult to predict at the time the token was issued. In CSA's assessment, NIST IR 8587's emphasis on short token lifetimes, rapid revocation, and continuous monitoring is an appropriate response to this dynamic, but it depends on organizations first establishing the non-human identity governance layer, discovery, ownership, and lifecycle tracking for every credential, that most enterprises have not yet built.

The post-quantum cryptography considerations in NIST IR 8587 deserve separate attention because they intersect with the same governance gap from a different angle. Migrating to quantum-resistant signature schemes will change the practical size and performance characteristics of tokens across an organization's identity infrastructure, and agencies that have not yet inventoried where and how tokens are issued, validated, and cached will be poorly positioned to assess where those larger keys and signatures will cause friction, whether in browser cookie size limits, HTTP header constraints, or throughput on hardware security modules tuned for classical algorithms [5]. In effect, the same discovery and inventory work that non-human identity governance requires today is also the prerequisite work for a smooth post-quantum migration tomorrow, which strengthens the case for treating token and identity inventory as foundational infrastructure rather than a lower-priority hygiene task.

Recommendations

Immediate Actions

Agencies and cloud service providers should begin an inventory of every system that issues or verifies identity tokens and assertions, mapping each signing key to the systems and identities it authorizes, since NIST IR 8587's key protection and scoping recommendations cannot be implemented against an unknown attack surface. Organizations operating AI agents should identify every agent-held credential currently in use, flag any credential shared across multiple agent instances, and prioritize retiring the shared-key pattern in favor of per-agent or per-session credentials with narrower scope. Security teams should also review token verification logic specifically, not just token issuance, given that the Storm-0558 incident succeeded in part because of a validation gap rather than solely because of the stolen key.

Short-Term Mitigations

Organizations should implement short-lived tokens with automated rotation for both human and non-human identities, targeting the hours-to-weeks credential lifetimes that NIST IR 8587 recommends, rather than the months-to-years lifetimes that remain common in many service-account and agent deployments based on industry reporting on credential rotation practices [5][6]. Continuous monitoring for anomalous token use, including unexpected geographic origin, unusual scope escalation, or token reuse patterns inconsistent with an agent's normal operating behavior, should be established where it does not already exist. Agencies and cloud providers should also begin documenting a formal AI agent identity lifecycle policy that specifies how agent identities are created, justified, reviewed, and decommissioned, directly addressing the governance vacuum that current survey data indicates most organizations have not yet filled [7].

Strategic Considerations

Security leadership should treat non-human identity governance as a prerequisite for, not a parallel track to, NIST IR 8587 compliance, since token-level protections are only as effective as the inventory and accountability structure that determines which tokens should exist in the first place. Organizations beginning post-quantum migration planning should sequence identity and token infrastructure inventory work early, since the same discovery process needed for non-human identity governance also surfaces where larger post-quantum keys and signatures will strain existing systems. Finally, organizations extending zero trust architectures to cover AI agents should treat agents as first-class principals subject to continuous verification and dynamic, least-privilege access, consistent with the direction of both NIST IR 8587 and the zero trust frameworks CSA and others have published for agentic systems, rather than retrofitting human-oriented identity controls onto a fundamentally different class of identity.

CSA Resource Alignment

NIST IR 8587's extension of token protection to AI agent identity connects most directly to CSA's research note The Non-Human Identity Governance Vacuum: AI Agents and the Fastest-Growing Unmanaged Attack Surface [9], which independently arrived at the conclusion that non-human identity governance, not any single technical control, is a primary security gap of the agentic AI era, and which recommends the same short credential lifetimes, centralized credential registries, and rapid revocation targets that NIST IR 8587 specifies for federal token infrastructure. Read together, the two documents

show a federal standards body and independent industry research converging on the same prescription from different starting points: NIST from a token-forgery incident response, and CSA from a broader survey of non-human identity sprawl.

CSA's artifact Agentic AI Identity and Access Management: A New Approach [10] extends this analysis by arguing that traditional protocols such as OAuth 2.1, SAML, and OpenID Connect, the same protocol family NIST IR 8587 assumes as its baseline, are not sufficient on their own for autonomous multi-agent systems, and proposes replacing long-lived bearer tokens with just-in-time, cryptographically signed attestations tied to a persistent agent identifier. This proposal is a more aggressive version of the token-lifetime and scoping recommendations in NIST IR 8587 and gives agencies a concrete architectural direction for the AI agent provisions the federal guidance introduces only at a high level.

For organizations seeking a broader protocol reference to implement NIST IR 8587's single sign-on, federation, and API protection recommendations, CSA's Navigating Identity and Access Management (IAM) [11] provides a comprehensive guide to the authentication, authorization, and provisioning standards underlying token-based access, covering both human and non-human identity scenarios. Finally, organizations mapping NIST IR 8587's requirements into an enterprise control framework should reference CSA's AI Controls Matrix (AICM) v1.1 AI Controls Matrix [12], particularly its identity and access management domain, which extends CSA's Cloud Controls Matrix baseline to account for the AI-specific and agentic identity risks that NIST IR 8587 addresses only for the federal and cloud-provider context.

References

[1] NIST. "NIST Finalizes Guidelines on Protecting Online Identity and Access Tokens From Misuse." NIST, September 2026.

[2] Help Net Security. "NIST and CISA finalize playbook to stop token theft and forgery." Help Net Security, September 16, 2026.

[3] NIST. "NIST Interagency Report 8587: Protecting Tokens and Assertions from Forgery, Theft, and Misuse." NIST Computer Security Resource Center, September 15, 2026.

[4] Help Net Security. "A "cascade" of errors let Chinese hackers into US government inboxes." Help Net Security, April 3, 2024.

[5] cybersecuritynews.com. "CISA and NIST Releases Technical Checklist for Safeguarding the Identity Tokens From Theft and Misuse." Cyber Security News, September 2026.

[6] IANS Research. "AI Agents Are Creating an Identity Security Crisis in 2026." IANS Research, April 19, 2026.

[7] Cloud Security Alliance. "79% of IT Pros Feel Ill-Equipped to Prevent Attacks Via Non-Human Identities, Cloud Security Alliance and Oasis Security Survey Finds." Cloud Security Alliance, January 27, 2026.

[8] Cloud Security Alliance. "Shadow AI and the Enterprise Visibility Crisis." CSA AI Safety Initiative, May 28, 2026.

[9] Cloud Security Alliance. "The Non-Human Identity Governance Vacuum: AI Agents and the Fastest-Growing Unmanaged Attack Surface." CSA AI Safety Initiative, May 20, 2026.

[10] Cloud Security Alliance. "Agentic AI Identity and Access Management: A New Approach." Cloud Security Alliance, August 18, 2025.

[11] Cloud Security Alliance. "Navigating Identity and Access Management (IAM)." Cloud Security Alliance, 2026.

[12] Cloud Security Alliance. "AI Controls Matrix (AICM) v1.1." Cloud Security Alliance, 2026.

