Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: 20 September 2026

J. Williams
Independent

19 March 2026

The Intent Token: A Cryptographic Authorization Primitive for Autonomous Agents
draft-williams-intent-token-00

Abstract

This document specifies the Intent Token, a cryptographic authorization primitive for autonomous AI agent systems. An Intent Token binds an autonomous agent action to a cryptographically signed, human-declared authorization envelope before that action is executed. The Intent Token addresses a fundamental gap in existing authorization frameworks: while OAuth 2.0, OIDC, and related standards govern identity and access at the session level, no standardized primitive exists for governing what an autonomous agent is authorized to DO at the moment of action. The Intent Token provides this primitive. It is model-agnostic, transport-agnostic, and composable with existing authorization infrastructure.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 20 September 2026.

Copyright Notice

Table of Contents

1. Introduction 2
2. Terminology 4
3. Problem Statement 4
3.1. The Authorization Gap Beyond OAuth 2.0 5
3.2. The Delegation Depth Problem 5
3.3. The Confused Deputy Problem in Autonomous Systems 5
4. The Intent Token 5
4.1. Token Structure 5
4.2. Required Claims 6
4.3. Temporal Scope and Shard Binding 7
4.4. Delegation Chain Encoding 7
5. Intent Certificate 7
6. Enforcement Binding 8
6.1. Pre-Action Validation 8
6.2. SNAP-BACK on Violation 8
6.3. Audit Chain Requirements 8
7. Interoperability 8
7.1. Relationship to OAuth 2.0 8
7.2. Relationship to Delegation Capability Tokens 9
7.3. Relationship to W3C Verifiable Credentials 9
8. Security Considerations 9
9. IANA Considerations 9
10. References 10
10.1. Normative References 10
10.2. Informative References 10
Author's Address 11

1. Introduction

The deployment of autonomous AI agents in consequential domains -- financial services, healthcare, critical infrastructure, autonomous vehicles, and multi-agent orchestration systems -- creates a class of authorization problems that existing standards do not address.

OAuth 2.0 [RFC6749] governs whether a principal has access to a resource. OpenID Connect [OIDC] governs whether a principal is who they claim to be. These standards address the WHO question: who is this agent, and what resources may it access?

Neither standard addresses the WHAT question: given that an agent has access, what specific actions is it authorized to perform at this moment, on behalf of this principal, within these declared boundaries?

The gap between access and action is the authorization gap. In human-operated systems, this gap is bridged by human judgment. In autonomous agent systems, this gap is bridged by the agent itself -- which may interpret its authorization scope in ways the authorizing principal did not intend. This is the intent blindness problem.

This document specifies the Intent Token, a cryptographic primitive that closes the authorization gap. An Intent Token is a signed, time-bounded authorization envelope that:

* Declares what action a principal authorizes an agent to perform, BEFORE the action is executed;

* Binds that declaration cryptographically to the agent session and the specific action;

* Provides a verifiable, tamper-evident record of the authorization for audit purposes;

* Propagates the declared intent through delegation chains, preventing sub-agents from exceeding the scope the delegating principal was itself authorized to grant.

The Intent Token is the core primitive of the Intent Bound Authorization (IBA) framework, patent application GB2603013.0 (pending, UK IPO, filed February 5, 2026, PCT rights in 150+ countries until August 2028).

Independent convergence on this primitive was identified in Google DeepMind’s introduction of Delegation Capability Tokens (DCTs) [DEEPMIND-DCT], February 12, 2026, seven days after the IBA patent filing. DeepMind’s paper identifies the same open problem: no standardized ontology for intent and responsibility across autonomous agent platforms exists. This document proposes the Intent Token as the candidate standard to fill that gap.

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Terminology

Intent Token: A cryptographically signed authorization envelope that declares what action a principal authorizes an autonomous agent to perform, issued before the action is executed, and bound to a specific session shard.

Intent Certificate: A persistent, verifiable credential that records the full authorization context for an agent session, including the principal identity, declared scope, temporal bounds, and cryptographic proof of all enforcement decisions made during the session.

Principal: The human or organizational entity whose declared intent authorizes the agent action. The principal MUST be cryptographically identified in every Intent Token.

Authorized Agent: An autonomous system authorized to act within the scope declared in an Intent Token.

Shard: A time-bounded cryptographic token issued by the IBA Gate that must be presented with every agent action. Shards expire after a configurable window (default 512 seconds).

IBA Gate: The enforcement point that validates Intent Tokens before permitting agent actions.

SNAP-BACK: The mandatory cancellation of an agent action when Intent Token validation fails. SNAP-BACK MUST occur before the action has market or physical effect.

Delegation Chain: The ordered sequence of principals and agents through which authorization has been delegated. Each link is cryptographically signed. No link may grant scope exceeding the scope granted to the delegating entity.

TBDE: Temporal Boundary Decision Engine. The runtime component responsible for validating every agent action against the declared Intent Token within the enforcement latency target.

WitnessBound Audit Chain: An append-only, cryptographically chained audit record of all enforcement decisions providing tamper-evident proof of all authorization decisions.

3. Problem Statement

3.1. The Authorization Gap Beyond OAuth 2.0

OAuth 2.0 access tokens grant access to resource scopes. A token with scope "read:financial_data" grants read access to financial data. It does not specify which financial data, under what conditions, at what time, or within what value limits.

An autonomous agent with an access token has no cryptographic constraint on what it does within its granted scope -- including actions the authorizing principal never contemplated. Existing mitigations -- rate limiting, anomaly detection, audit logging -- are post-hoc. The Intent Token is pre-hoc: it declares and enforces authorization before action.

3.2. The Delegation Depth Problem

Modern autonomous agent architectures are hierarchical. An orchestrating agent delegates tasks to sub-agents, which may delegate further. Existing delegation frameworks, including OAuth 2.0 Token Exchange [RFC8693], do not provide a mechanism for propagating declared human intent through the delegation chain.

The Intent Token addresses this by requiring that every link in the delegation chain present a valid Intent Token that is a subset of the authorizing token. A sub-agent cannot act outside the scope its delegating principal was itself authorized to grant.

3.3. The Confused Deputy Problem in Autonomous Systems

The confused deputy problem [HARDY-1988] describes a scenario in which a system with legitimate authority is manipulated into misusing that authority. In autonomous agent systems, an agent may be manipulated via prompt injection, adversarial input, or environmental manipulation into taking actions outside its intended scope.

The Intent Token defends against confused deputy attacks. Because the authorized action is declared and signed by the principal BEFORE the agent acts, any deviation from the declared intent -- regardless of cause -- is detected and blocked at the enforcement layer.

4. The Intent Token

4.1. Token Structure

An Intent Token is a JSON Web Token [RFC7519] with the following structure. The header MUST include "typ": "intent+jwt".

```
{
  "header": {
    "alg": "ES256",
    "typ": "intent+jwt"
  },
  "payload": {
    "iss": "https://intentbound.example.com",
    "sub": "agent-id:AG-7729-ALPHA",
    "aud": "iba-gate:production",
    "exp": 1771234567,
    "iat": 1771230967,
    "jti": "a8f3c2d1-4b5e-6f7a-8b9c-0d1e2f3a4b5c",
    "ibt_ver": "1.0",
    "principal": {
      "id": "user:jeffrey.williams@example.com",
      "type": "human",
      "auth_method": "ES256-keypair"
    },
    "declared_intent": {
      "action_class": "financial:order:equity",
      "scope": "EQUITIES_US",
      "bounds": {
        "max_position_usd": 50000000,
        "max_single_order_usd": 5000000,
        "permitted_instruments": ["NYSE:*", "NASDAQ:*"],
        "prohibited_instruments": ["OTC:*"],
        "time_window": "09:30-16:00 EST"
      }
    },
    "shard": {
      "id": "shard:mmc-nbwc-enyn",
      "issued_at": 1771230967,
      "expires_at": 1771231479,
      "window_seconds": 512
    },
    "enforcement": {
      "snap_back": true,
      "latency_target_ms": 5,
      "audit_chain": "witnessbound:production"
    }
  }
}
```


4.2. Required Claims

The following claims are REQUIRED in every Intent Token:

ibt_ver: The version of the Intent Token specification. This

document defines version "1.0".

principal: An object identifying the authorizing human principal. MUST include "id" (a URI) and "type" ("human" or "organization").

declared_intent: An object declaring the specific action class the principal authorizes. MUST include "action_class" and "scope".

shard: An object binding the token to a specific time-bounded session shard. MUST include "id", "issued_at", "expires_at", and "window_seconds".

enforcement: An object specifying enforcement requirements. MUST include "snap_back" (boolean, MUST be true) and "audit_chain" (URI of the audit chain endpoint).

Standard JWT claims iss, sub, aud, exp, iat, and jti are REQUIRED.

4.3. Temporal Scope and Shard Binding

Every Intent Token MUST be bound to a time-bounded shard. The shard provides temporal scope and session binding. The default shard window is 512 seconds. When a shard expires, the agent MUST obtain a new Intent Token before continuing to act. The enforcement layer MUST reject any action presented with an expired shard.

4.4. Delegation Chain Encoding

When an authorizing agent delegates to a sub-agent, the delegating agent MUST issue a new Intent Token with a declared_intent scope that is a subset of its own authorized scope. A delegating agent MUST NOT grant scope exceeding its own authorization. The delegation_chain array MUST include all prior delegation records plus a new signed record for this delegation.

5. Intent Certificate

An Intent Certificate is the persistent audit record of an agent session. It MUST contain: the principal identity and authentication record; the declared scope for the session; the temporal bounds; a reference to the WitnessBound audit chain; cryptographic hashes of all Intent Tokens issued during the session; and a signed summary of all enforcement decisions including SNAP-BACK events.

Intent Tokens and Intent Certificates MUST be signed using algorithms from NIST SP 800-57 [NIST-SP-800-57]. Implementations MUST support ES256 [RFC7518]. Implementations SHOULD support ES384 and EdDSA [RFC8037] for higher-security deployments.

6. Enforcement Binding

6.1. Pre-Action Validation

The IBA Gate MUST validate the Intent Token BEFORE the agent action is permitted to execute. The IBA Gate MUST verify: token signature validity; token and shard expiry; action matches declared_intent action_class; action parameters are within declared bounds; delegation chain validity; and jti uniqueness for replay prevention.

Implementations MUST complete all validations within the latency_target_ms specified in the enforcement claim. The target is 5ms.

6.2. SNAP-BACK on Violation

If any validation check fails, the IBA Gate MUST: reject the agent action immediately; record a SNAP-BACK event in the WitnessBound audit chain; return a structured error to the agent; and NOT permit the action to be retried without a new Intent Token authorized by the principal. SNAP-BACK is a hard stop. No partial execution is permitted.

6.3. Audit Chain Requirements

Every enforcement decision MUST be recorded in the WitnessBound audit chain. Each record MUST contain: the jti of the Intent Token; the validation timestamp (nanosecond precision RECOMMENDED); the validation result (PASS or SNAP-BACK); and a cryptographic hash chained to the previous record. This audit chain satisfies the audit requirements of EU AI Act Article 9, MiFID II, FDA Cybersecurity 2023, and HIPAA Security Rule.

7. Interoperability

7.1. Relationship to OAuth 2.0

The Intent Token is complementary to OAuth 2.0 [RFC6749]. OAuth 2.0 governs access; the Intent Token governs action. In a compliant implementation: the agent obtains an OAuth 2.0 access token establishing resource access; the principal issues an Intent Token declaring the specific authorized action; the IBA Gate validates the Intent Token before permitting the action.

7.2. Relationship to Delegation Capability Tokens

Tomasev et al. [DEEPMIND-DCT] introduced Delegation Capability Tokens (DCTs) as a primitive for AI agent authorization. DCTs address the same authorization gap as the Intent Token, arriving independently seven days after the IBA patent filing. The IETF is encouraged to consider harmonization between this specification and the DCT work, given the independent convergence on the same primitive.

7.3. Relationship to W3C Verifiable Credentials

Intent Certificates MAY be implemented as W3C Verifiable Credentials [VC-DATA-MODEL], using the "IntentCertificate" credential type with credential subject claims encoding the session authorization context specified in Section 5.

8. Security Considerations

Replay Attacks: Intent Tokens MUST include a unique jti claim. The IBA Gate MUST maintain a record of used jti values for the duration of the shard window.

Token Forgery: Implementations MUST follow NIST SP 800-57 for key management. Hardware security modules (HSMs) are RECOMMENDED for principal signing keys in high-security deployments.

Scope Escalation: Implementations MUST validate the complete delegation chain, not only the most recent delegation record.

Prompt Injection: The Intent Token provides defense against prompt injection attacks. An agent manipulated into attempting an out-of-scope action will be stopped by SNAP-BACK regardless of the cause.

Enforcement Latency: The 5ms enforcement latency target is a security parameter. In financial systems, actions with market impact occur faster than 5ms. Implementations in latency-sensitive domains MUST measure and report enforcement latency.

Audit Chain Integrity: Implementations MUST store audit chain records in append-only storage with access controls preventing modification or deletion.

9. IANA Considerations

This document requests the following JWT Claim registrations:

* Claim Name: ibt_ver -- Intent Token specification version

* Claim Name: declared_intent -- Principal-declared authorized action scope

* Claim Name: shard -- Time-bounded session shard binding

* Claim Name: enforcement -- Enforcement requirements for this token

* Claim Name: delegation_chain -- Ordered delegation records

This document requests registration of the media type application/intent+jwt for Intent Tokens.

10. References

10.1. Normative References

[NIST-SP-800-57] National Institute of Standards and Technology, "Recommendation for Key Management: Part 1 - General", NIST Special Publication 800-57 Part 1 Rev 5, May 2020.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC6749] Hardt, D., Ed., "The OAuth 2.0 Authorization Framework", RFC 6749, October 2012, <https://www.rfc-editor.org/info/rfc6749>.

[RFC7518] Jones, M., "JSON Web Algorithms (JWA)", RFC 7518, May 2015, <https://www.rfc-editor.org/info/rfc7518>.

[RFC7519] Jones, M., Bradley, J., and N. Sakimura, "JSON Web Token (JWT)", RFC 7519, May 2015, <https://www.rfc-editor.org/info/rfc7519>.

[RFC8037] Liusvaara, I., "CFRG Elliptic Curves for JOSE", RFC 8037, January 2017, <https://www.rfc-editor.org/info/rfc8037>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC8693] Jones, M., Nadalin, A., Campbell, B., Ed., Bradley, J., and C. Mortimore, "OAuth 2.0 Token Exchange", RFC 8693, January 2020, <https://www.rfc-editor.org/info/rfc8693>.

10.2. Informative References

[DEEPMIND-DCT]
Tomasev, N., "Delegation Capability Tokens for AI Agent Authorization", arXiv 2602.11865, February 2026.

[HARDY-1988]
Hardy, N., "The Confused Deputy (or why capabilities might have been invented)", SIGOPS Operating Systems Review Volume 22, Issue 4, October 1988.

[OIDC] Sakimura, N., Bradley, J., Jones, M., de Medeiros, B., and C. Mortimore, "OpenID Connect Core 1.0", November 2014.

[VC-DATA-MODEL]
Sporny, M., "Verifiable Credentials Data Model v2.0", May 2024.

Author's Address

Jeffrey Williams
Independent
Chiang Mai
Thailand
Email: urls@live.com