RATS
Internet-Draft
Intended status: Standards Track
Expires: 15 December 2025

K. Huang
DistributedApps.ai

June 2025

Capability Attestation Extensions for the Entity Attestation Token (EAT) in Agentic AI Systems

draft-huang-rats-agentic-eat-cap-attest-00

Abstract

This document specifies extensions to the Entity Attestation Token (EAT) [RFC9248] to support robust, interoperable attestation of capabilities in agentic AI systems. These extensions introduce new claims and guidance for securely asserting agent functional, reasoning, and operational capabilities, as well as their compositional structure and policy constraints. The goal is to enable trustworthy, verifiable, and privacy-respecting capability attestation for autonomous agents in dynamic, decentralized environments.

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 3 December 2025.

Copyright Notice

Table of Contents

1. Introduction 2
2. Terminology 2
3. Use Cases 3
4. Capability Attestation Claims 3
5. Nested and Modular Agent Representations 4
6. Endorsements and Trust Anchors 5
7. Security Considerations 5
8. Privacy Considerations 5
9. IANA Considerations 6
10. Normative References 6
11. Informative References 6
Author's Address 6

1. Introduction

The Entity Attestation Token (EAT) [RFC9248] defines a CBOR/COSE-based structure for representing signed claims about an entity's identity, configuration, and operational state. While EAT is widely adopted for device attestation, agentic AI systems— such as autonomous planners, LLM-based agents, and API orchestrators— require more granular and dynamic attestation of their capabilities, constraints, and compositional structure.

This document defines EAT extensions for agentic AI, supporting:

* Attestation of agent capabilities, behavioral policies, and internal module composition.

* Secure, privacy-preserving assertions suitable for decentralized, multi-agent environments.

* Mechanisms for endorsement and trust anchor management.

These extensions are intended to facilitate secure agent interaction, policy-based access control, and dynamic trust establishment.

2. Terminology

Agent An autonomous computational entity capable of executing plans, interacting with services, and making decisions.

Capability Attestation The process of proving what an agent can do, including reasoning methods, tool use, language models, and planning systems.

Agent Capability Token (ACT) An EAT-compliant token carrying claims about agentic capability attributes.

Submodule A functional component within an agent (e.g., planner, retriever, executor) that may have its own EAT claims.

3. Use Cases

* Trust establishment between AI agents prior to interaction, ensuring only agents with appropriate capabilities participate in sensitive workflows.

* Secure registration of AI agents in public or private registries, with verifiable claims about their operational scope and limitations.

* Policy-based access control where access to resources or APIs is granted based on attested agent capabilities and policy constraints.

* Dynamic capability negotiation in multi-agent systems, enabling agents to adaptively select partners or workflows based on verified capabilities.

4. Capability Attestation Claims

The following claims are introduced for agent capability attestation. Each claim is assigned a unique CBOR label in the EAT claims registry.

agent_id (CBOR label 40001) Globally unique identifier for the agent.

agent_capabilities (CBOR label 40002) Map describing capabilities, e.g., planning methods, NLP models, tool use, reasoning, delegation.

policy_constraints (CBOR label 40003) Operational policies and constraints, e.g., data access, temperature limits, explainability.

capability_version (CBOR label 40004) Version string of the capability declaration.

model_fingerprint (CBOR label 40005) Hash or identifier of the core model or weights.

dynamic_proof (CBOR label 40006) Challenge-response or external validation artifact.

submodules (CBOR label 40007) Array of nested EATs, each representing a submodule with its own signed claims.

endorsements (CBOR label 40008) Endorsement by registry or certifying authority, including issuer, cert_type, and signature.

Example agent_capabilities claim:

```
{
  "planning": ["BFS", "A*", "LlamaPlan"],
  "nlp_models": ["llama3-8b", "gpt-4.5-turbo"],
  "tool_use": ["web_access", "code_exec"],
  "reasoning": ["symbolic", "LLM-hybrid"],
  "delegation": true
}
```

Example policy_constraints claim:

```
{
  "data_access": ["PII_restricted"],
  "temperature_limit": 0.8,
  "explainability_required": true
}
```

5. Nested and Modular Agent Representations

Agentic AI systems may be composed of multiple modules, each with distinct capabilities and trust requirements. The submodules claim enables the inclusion of multiple signed, nested EATs, each representing a submodule. Each submodule EAT must include its own agent_capabilities and be signed by the same or a recognized authority.

This compositional approach supports modular attestation, allowing verifiers to assess the trustworthiness of both the agent as a whole and its individual components.

6. Endorsements and Trust Anchors

Endorsements provide third-party assurance of agent capability claims. The endorsements claim encodes information such as the issuer, certificate type, and a COSE_Sign1 signature over the claims or schema.

Example endorsements claim:

```
{
  "issuer": "AgenticAITrust.org",
  "cert_type": "capability-schema",
  "signature": "<COSE_Sign1 representation>"
}
```

Trust anchors for capability validation should be managed by ecosystem authorities, using X.509 or DICE profiles as appropriate. Verifiers must validate endorsement signatures and check certificate revocation status as part of the trust evaluation process.

7. Security Considerations

* All claims must be signed using COSE_Sign1. Endorsements should be cryptographically verifiable.

* Include freshness indicators such as iat (issued at), exp (expiration), and nonces for replay protection.

* The dynamic_proof claim enables challenge-response or external validation to demonstrate live capability.

* Only disclose claims necessary for the verifier’s trust decision, minimizing exposure of internal details.

* Implementers must ensure compliance with relevant security best practices for cryptographic operations and key management.

8. Privacy Considerations

* Capability claims may reveal sensitive internal structure. Use COSE_Encrypt for confidentiality when required.

* Selective disclosure via layered EATs can support verifier-specific access.

* Implementers must ensure compliance with relevant privacy laws and regulations when attesting capabilities.

9. IANA Considerations

This document requests allocation of CBOR labels 4000140008 in the Entity Attestation Token (EAT) claims registry.

10. Normative References

[RFC9248] Lundblade, L., Mandyam, G., and J. O'Donoghue, "The Entity Attestation Token (EAT)", RFC 9248, June 2022, <https://www.rfc-editor.org/rfc/rfc9248.html>.

11. Informative References

[RFC9334] Birkholz, H., Thaler, D., Eckel, M., and N. Smith, "Remote Attestation Procedures Architecture", RFC 9334, January 2023, <https://www.rfc-editor.org/rfc/rfc9334.html>.

Author's Address

Ken Huang
DistributedApps.ai
Email: ken.huang@DistributedApps.ai