Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: 10 July 2026

M. Chen
L. Su
China Mobile

6 January 2026

New requirements for Authentication and Authorization in the AI Agents era

draft-chen-ai-agent-auth-new-requirements-00

Abstract

AI Agents are rapidly evolving from academic concepts into the core engines driving next-generation applications. However, their autonomy, dynamic nature, and complex delegation relationships pose a fundamental challenge to our existing authentication and authorization frameworks, which were designed for human users and traditional software. This document dissects the novel characteristics of AI Agents and outlines the new requirements for authentication and authorization which can manage dynamic behavior rather than verifying static identity.

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 10 July 2026.

Copyright Notice

Table of Contents

1. Introduction

Traditional security models are built on a core assumption: the behavior of a protected entity, be it a human or a service, is relatively predictable. We authenticate an "identity" and then grant it a set of fixed "permissions." AI Agents shatter this foundation.

An AI Agent is not a simple instruction executor; it is a goal achiever. We provide it with a high-level objective (e.g., "optimize supply chain costs"), and it autonomously deconstructs the task, learns from its environment, invokes tools, and may discover innovative operational paths we never anticipated. This shift introduces four disruptive characteristics:

High Autonomy and Emergent Behavior: An Agent's actions are not pre-coded but are dynamically generated to achieve a goal. Static permission rules can neither foresee nor cover all its possible operations.

Dynamic, Ephemeral, and Replicable Nature: For efficiency, a primary Agent may spawn thousands of ephemeral sub-agents in an instant to handle parallel tasks. Their identities are transient, massive in scale, and may exist for only milliseconds.

Complex Delegation and Chains of Responsibility: Agents can form deep, networked call chains. A travel Agent might call a flight Agent, which in turn calls a payment Agent. When an unauthorized action occurs, attributing responsibility and tracing the flow of permissions becomes incredibly complex.

Continuous Learning and Adaptation: An Agent's decision-making model evolves over time. This means an Agent's "normal behavior" today may differ from yesterday's, making it vulnerable to model drift or malicious manipulation that traditional static credentials cannot detect.

2. Terminology

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].

3. New Authentication Requirements:From "Who Are You?" to "Are You Still Trustworthy?"

In the face of these characteristics, one-time authentication becomes woefully inadequate. We need a new authentication framework capable of continuously assessing an Agent's trustworthiness.

3.1. New Requirement 1: Traceable Identity and Provenance

Authentication protocols must support a new identity format that is more than just an ID; it's a cryptographic "provenance record." This certificate must contain:

Genesis Information: The root user or task that initiated the Agent's creation. Delegation Chain: A cryptographically signed, non-forgeable call path that clearly records every delegation from the genesis to the current instance.

3.2. New Requirement 2: Continuous Behavioral Attestation

Authentication cannot be a single event. Protocols must support a lightweight "behavioral heartbeat" mechanism, allowing an Agent to periodically submit a cryptographic digest of its recent actions to a monitoring system. By comparing this digest against an expected behavioral baseline, the system can continuously verify that the Agent is "acting normally," thus detecting hijacking or unexpected drift even if its identity credential remains valid.

3.3. New Requirement 3: Ultra-Low-Overhead Identity Management

To support massive-scale, ephemeral Agents, protocols must enable near-zero overhead for identity creation and verification. Expensive public-key operations and complex handshakes should be replaced with efficient symmetric-key mechanisms to meet the demands of high-frequency creation, destruction, and continuous attestation.

4. New Authorization Requirements: From "What Permissions Do You Have?" to "What Are You Allowed to Do, Right Here, Right Now?"

Static, Role-Based Access Control (RBAC) is obsolete in the face of autonomous Agents. Authorization decisions must become dynamic, precise, and risk-aware.

4.1. New Requirement 1: Intent-Driven, Just-in-Time (JIT) Permissions

Authorization is no longer about pre-assigning a broad role. Instead, it must be about granting the principle of least privilege required to complete a specific task, at the moment of execution. For example, the system should grant an Agent permission to "execute a payment of amount <= $100 for the purpose of booking a flight" rather than a generic "payment" permission. This permission must expire immediately upon task completion.

4.2. New Requirement 2: Rich Context and Risk-Awareness

Authorization decisions must be based on a rich set of contextual factors, including but not limited to:

* The Agent’s intent and objective.
* A risk assessment of the requested operation.
* Whether its current behavioral patterns are normal.
* The sensitivity and provenance of the data being accessed.

Protocols must be able to efficiently carry this structured context to the policy engine for real-time evaluation.

4.3. New Requirement 3: Explainable and Auditable Authorization

Given the autonomy of Agents, when something goes wrong, we must be able to answer, "Why was the system authorized to do that?" Therefore, the response from an authorization protocol must be explainable. It should return not just an "Allow" or "Deny" verdict but also the rationale behind the decision, such as the policy IDs that were matched and a snapshot of the critical context evaluated. This is essential for post-mortem audits, accountability, and the iterative refinement of security guardrails.

5. IANA Considerations

6. Security Considerations

Authors' Addresses

Meiling Chen
China Mobile
BeiJing
China
Email: chenmeiling@chinamobile.com

Li Su
China Mobile
BeiJing
China
Email: suli@chinamobile.com