Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: 18 August 2026

M. Chen
L. Su
China Mobile

14 February 2026

A Decoupled Authorization Model for Agent2Agent
draft-chen-agent-decoupled-authorization-model-00

Abstract

This document proposes a framework for dynamic, intent-based authorization for AI Agents. The primary goal is to enable fine-grained, Just-in-Time (JIT) permissions based on an agent's specific intent and behavioral trustworthiness, rather than a long-lived identity or role, achieve decoupling of authorization policies from business operations.

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 18 August 2026.

Table of Contents

1. Introduction

AI Agents, hereafter "agents" represent a significant evolution from traditional scripts or services. They possess the ability to reason, plan, and execute multi-step tasks to achieve a high-level goal. This autonomy, however, creates a significant attack surface. An agent granted broad permissions, if compromised via mechanisms like prompt injection, can cause widespread damage.

Existing authorization models are ill-suited for this dynamic environment. They typically bind permissions to stable, long-lived identities. Agents, by contrast, are often ephemeral, created in large numbers for specific, short-lived tasks.

This document proposes a model that decouples authorization decisions from the services themselves. It leverages a Authorization Decision Point (ADP) that makes real-time, context-rich decisions for every action an agent attempts to take. The core of this draft is a standardized "Input Contract" -- a structured data format that a Authorization Execution Point (AEP) MUST provide to the ADP to enable fine-grained, intent-based authorization.

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].

* AI Agent (Agent): An autonomous or semi-autonomous software entity that perceives its environment and takes actions to achieve goals.

* Intent: A declaration of the specific goal or task an agent intends to accomplish. This is typically more granular than a broad scope.

* Ephemeral Identity: A short-lived, single-purpose identity assigned to an agent for the duration of its task.

* Authorization Decision Point (ADP): The logical component that evaluates policies and makes authorization decisions (e.g., Permit/Deny).

* Authorization Execution Point (AEP): The logical component that intercepts an agent's action, requests a decision from the ADP, and enforces that decision.

3. Problem Statement

Traditional authorization mechanisms, often designed for human-to-machine interactions, are ill-suited for the dynamic and large-scale nature of agent-to-agent communication. The key challenges are categorized as follows. Traditional authorization mechanisms fail to address the unique characteristics of AI Agents in several key areas.

Scale and Complexity Explosion: In a system with N agents and M resources, the number of potential authorization rules can grow combinatorially. Managing these rules through static methods like Access Control Lists (ACLs) becomes untenable.

Static Permissions vs Dynamic Intent: Assigning static roles or scopes (e.g., read:all, write:all) is overly permissive. An agent's authority should be scoped precisely to its immediate intent (e.g., "query flight from XX to XX under $500"), and this authority should be granted just-in-time.

Identity Lifecycle Mismatch: The high cost and administrative overhead of managing traditional identities (e.g., user accounts, service accounts) are incompatible with the massive scale and ephemeral nature of agents.

One-Time Trust vs Continuous Risk: An agent's behavior can be subverted at any point in its lifecycle. A successful authentication at the beginning of a session provides no guarantee of trustworthy behavior throughout. A continuous attestation of behavioral patterns is required.

Flat vs Hierarchical Authority: Agents often delegate tasks to sub-agents. A robust framework must support hierarchical delegation of authority, ensuring a sub-agent's permissions are a strict subset of its parent's, and providing a clear chain of accountability.

Governance and Auditability: When an undesirable action occurs, tracing the root cause across a chain of agent interactions is a formidable challenge.

4. Analysis of Oauth for Agent-to-Agent Authorization

The OAuth 2.0 Authorization Framework [RFC6749] and its related RFCs provide a strong foundation for secure, delegated access.

4.1. What OAuth2.0 can do

OAuth 2.0 effectively addresses the foundational aspects of secure delegation and credential management. It provides:

* A standard mechanism for Delegated Authorization, allowing agents to act on behalf of a user.

* A standard for Credential Passing (Tokens) crucial for interoperability.

* A grant type for Machine-to-Machine Authentication (Client Credentials).

* A mechanism for Limited-Scope and Time-Bound Access via scopes and token expiration.

4.2. Remaining Challenges and Open Issues

While indispensable, OAuth 2.0 is primarily a framework for delegation and token issuance. However, there are still some issues that need to be addressed on the AI agent internet.

Fine-Grained Policy Decision: The parameter "scope" in OAuth 2.0 is often too coarse for complex authorization logic. A resource server agent receiving a token with a read:data scope still needs to answer more detailed, context-specific questions (e.g., "Which specific record?", "Under what conditions?").

Unified Policy Language and Governance: OAuth 2.0 does not define a language for expressing authorization policies or a methodology for managing them. The interpretation of a scope is left to each agent's implementation. This makes it difficult to centrally manage, audit, or reason about the authorization policies across the entire system.

5. Decoupled Authorization Model

To address the open issues, this document proposes a Decoupled Authorization Model. This model complements OAuth 2.0 by externalizing the fine-grained policy decision logic from the agent's service-specific business logic.

5.1. Architecture overview

The model is based on the separation of the Authorization Decision Point (ADP) from the Authorization Execution Point (AEP). This decouples the "what" (the business logic) from the "if" (the authorization logic).

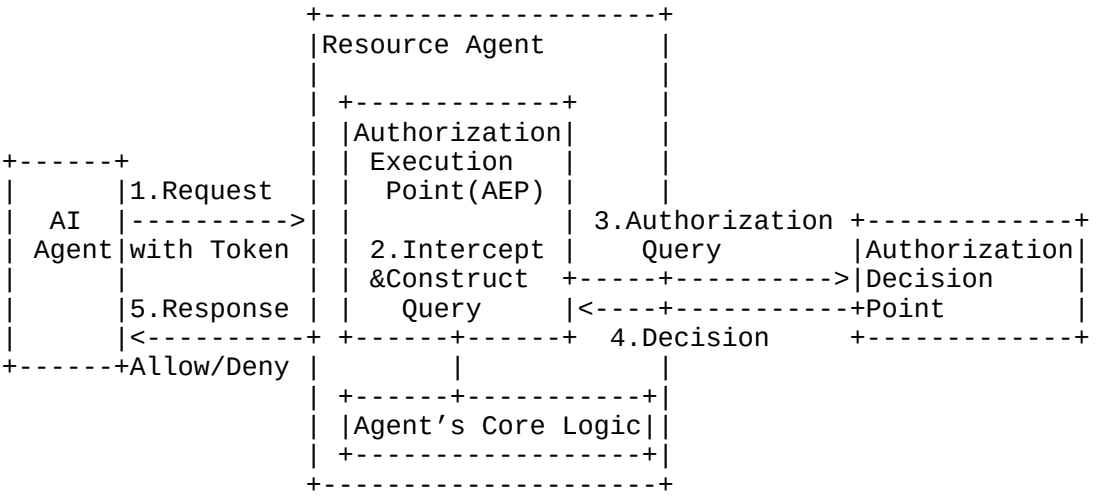

Figure 1: Decoupled Authorization Model Interaction Flow

5.2. Authorization Execution Point(AEP)

The AEP is a logical gateway protecting an agent's capabilities, responsible for constructing the complete context for the decision:

* Intercept an incoming request (Step 2).
* Perform initial validation of credentials (e.g., a JWT access token).
* Assemble a query for the ADP, containing all relevant context (e.g., token claims, request details).

* Send the query to the ADP (Step 3) and receive a simple Allow or Deny decision (Step 4).

* Enforce the decision by either forwarding the request to the agent’s core logic or rejecting it (Step 5).

5.3. Authorization Decision Point(ADP)

The ADP is the authorization "brain", responsible for evaluating the context against a set of policies:
- Receive authorization queries from the AEP. - Maintain a current set of authorization policies written in an expressive, declarative language. - Evaluate the query against the policies and data to produce a definitive decision. - The ADP SHOULD be deployable in a decentralized manner to ensure low latency and high availability.

5.4. Input

TBD

6. Security Considerations

TBD

7. IANA Considerations

TBD

8. Informative References

[RFC6749] Hardt, D., Ed., "The OAuth 2.0 Authorization Framework", RFC 6749, DOI 10.17487/RFC6749, October 2012, <https://www.rfc-editor.org/rfc/rfc6749>.

Authors’ Addresses

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

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