Network Working Group
Internet-Draft
Intended status: Informational
Expires: 3 October 2026

Y. Jiang
L. Li
Y. Song
F. Liu
Huawei

1 April 2026

Security Considerations for Intent-Based Requests in Agentic Systems
draft-jiang-intent-security-00

Abstract

Intent-based requests enable users, applications, and agents to express goals and constraints without specifying step-by-step procedures. Such intents are commonly translated into executable directives and propagated across multiple entities (clients, agents, authorization components, orchestration functions, and execution endpoints). This multi-hop processing expands the attack surface for tampering, privilege escalation, constraint bypass, and intent drift.

This document provides a solution-agnostic security analysis for intent-based requests across agentic systems. It introduces a reference model and scenarios to guide protocol and system design, and also presents threats and requirements. The document emphasizes constraint validation, invocation validation, multi-hop chain-of-custody, and policy-driven responses to drift, while remaining independent of any specific deployment domain.

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 October 2026.

Table of Contents

1. Introduction 3
1.1. Scope 3
2. Terminology and Conventions 3
2.1. Conventions 3
2.2. Definitions 3
2.3. Acronyms 4
3. Problem Statement and Threat Model 4
3.1. Threats 5
3.2. Requirements 5
4. Reference Model 6
4.1. Reference Model Entities 7
4.2. Operational Overview 7
5. Security Scenarios 8
5.1. Scenario 1: Directive Tampering and Authorization Boundary Expansion 8
6. Security Considerations 9
6.1. Scenario-to-Requirement Mapping 9
6.2. Considerations for Scenario 1 (Directive Tampering) . . . 10
6.2.1. Overview 10
6.2.2. Illustrative Procedure 10
6.3. General Security Considerations 11
7. IANA Considerations 12
8. Normative References 12
9. Informative References 12
Acknowledgments 13
Authors' Addresses 13

1. Introduction

Intent-based interaction is increasingly adopted in automation, orchestration, and agentic systems, where a request expresses desired outcomes and constraints rather than explicit procedures. A receiving system (or a chain of systems) translates the intent into executable directives and invokes tools or services to achieve the intended outcome.

Multi-hop processing (client-to-agent, agent-to-agent, agent-to-tool/service) introduces security risks beyond traditional single-hop APIs, including: (1) integrity and substitution attacks against derived directives, (2) privilege escalation during tool/service invocation, (3) constraint bypass, and (4) multi-hop intent drift where constraints degrade or diverge over transformations.

This document does not define a new protocol. Instead, it provides a security-oriented reference model, threat analysis, requirements, and scenarios to support future standardization and interoperable designs.

1.1. Scope

This document focuses on security considerations for intent-based requests in multi-hop agentic systems. While examples may reference telecom or networking contexts, the analysis applies broadly to any domain where intent processing spans multiple trust boundaries.

2. Terminology and Conventions

2.1. Conventions

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.

2.2. Definitions

This document uses the following terms:

Intent: A declarative expression of desired operational goals and outcomes, without specifying how to achieve or implement them. This definition is aligned with intent-based networking (IBN) guidance [RFC9315] [RFC9316].

Intent Translation: The process of transforming an intent into more

concrete representations, such as constraints, objectives, candidate procedures, or executable directives.

Constraint: A condition that limits acceptable outcomes or actions. Constraints may include invariants, policy rules, safety boundaries, and compliance requirements.

Constraint Validation: Verifying whether an intent and/or its derived artifacts comply with applicable constraints, invariants, policy rules, and safety boundary requirements.

Invocation: A request to a tool or service intended to fulfill an intent (e.g., API call, workflow step, actuation command).

Invocation Validation: Determining whether an invoker holds the required privileges to invoke a tool or service and whether invocation parameters satisfy the requirements and constraints specified by the intent.

Observation: Telemetry, events, measurements, or other signals used for monitoring and assurance.

Drift: A divergence between the intent (including its constraints) and the realized plan or actions over time or across multi-hop transformations.

Derived Directive: An executable or enforceable artifact generated from an intent through translation, such as an allowed rule set, capability token, or authorization grant.

2.3. Acronyms

IBN: Intent-Based Networking

IBS: Intent-Based System

3. Problem Statement and Threat Model

In many agentic systems, an intent is translated into executable directives (e.g., an allowed rule set) that must be propagated across multiple entities and enforced at execution endpoints. However, existing designs often lack end-to-end mechanisms that jointly ensure: (1) directives remain within authorized boundaries across transformations and propagation, (2) constraints are validated before execution, (3) invocations are privilege-checked and constraint-checked at each call boundary, and (4) drift is detected and handled under policy.

3.1. Threats

This document considers the following representative threats in multi-hop intent processing:

T1 (Directive Tampering/Substitution): A malicious or compromised intermediary modifies or replaces derived directives (e.g., allowed rule sets) to expand privileges or bypass constraints.

T2 (Unauthorized Invocation / Privilege Escalation): An agent or client invokes tools/services without the required privileges, or smuggles parameters that violate intent constraints.

T3 (Constraint Bypass): Constraints, invariants, or safety boundaries are dropped, weakened, or misapplied during translation, planning, or execution.

T4 (Multi-Hop Drift): Across repeated transformations, delegations, and tool invocations, the realized actions diverge from the original intent boundary (accidentally or adversarially).

T5 (Monitoring Evasion / False Observations): Attackers evade detection by suppressing, forging, or selectively presenting observations used for assurance and drift detection.

3.2. Requirements

Based on the threats above, this document identifies the following security requirements:

R1 (Provenance and Authorization Boundary Binding): The system provides a verifiable binding between the intent, derived directives, and the applicable authorization boundary, such that unauthorized expansion can be detected or prevented.

R2 (Chain-of-Custody for Derived Directives): The system protects derived directives against tampering and substitution across multi-hop propagation.

R3 (Constraint Validation): The system validates the intent and/or derived artifacts against applicable constraints, invariants, policy rules, and safety boundaries before accepting or executing actions.

R4 (Invocation Validation): The system validates that an invoker holds the required privileges to invoke a tool/service and that invocation parameters satisfy intent constraints prior to and/or at invocation time.

R5 (Non-Bypass Enforcement): The execution endpoint enforces constraints and authorization boundaries such that direct/side-path invocation cannot bypass required checks.

R6 (Observability and Auditability): The system provides sufficient observations and audit evidence to support compliance assessment, drift detection, and incident investigation.

R7 (Policy-Driven Drift Response): Upon drift detection or constraint violation, the system supports policy-driven responses (e.g., deny, degrade, re-confirm, re-negotiate, fallback).

4. Reference Model

This section introduces a technology-neutral reference model for intent-based requests. The model is aligned with intent-based system decomposition commonly used in IBN guidance [RFC9315], while remaining applicable to non-networking domains.

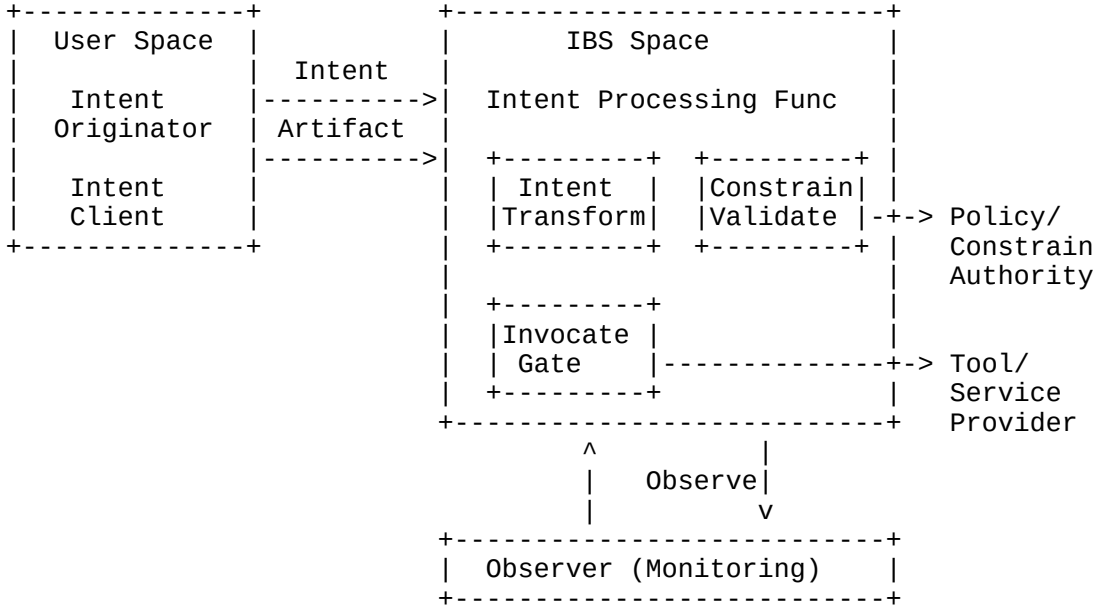

Figure 1: Reference Model for Multi-Hop Intent Processing

The figure separates User Space from IBS Space for clarity. Deployments may collapse functions into fewer components or distribute them across multiple agents and services.

4.1. Reference Model Entities

The following entities are defined in the reference model:

Intent Originator: The party whose goals and constraints are to be satisfied (e.g., human user, application owner, operator, or delegated principal).

Intent Client: The component that submits intents to an IBS and may carry contextual signals.

Intent Processing Function: A logical function that performs translation, validation, and orchestration for intent fulfillment. This function encompasses the Intent Transformer, Constraint Validator, and Invocation Gate.

Intent Transformer: A function that transforms intent representations (e.g., natural language to structured intent, structured intent to constraints/objectives, objectives to derived directives).

Constraint Validator: A function that enforces R3 by validating intents and derived artifacts against constraints, invariants, policy rules, and safety boundaries.

Invocation Gate: A function that enforces R4 and R5 by privilege-checking and constraint-checking each tool/service invocation and preventing bypass of required checks.

Policy/Constraint Authority: A logical source of constraints and policy boundaries (e.g., organizational policy, compliance rules, safety invariants, subscription/contract limits).

Tool/Service Provider: A system that executes actions (APIs, workflows, actuators, management functions, data services).

Observer (Monitoring Function): A function that collects observations (telemetry, events, measurements) used for assurance, compliance assessment, and drift detection (R6 and R7).

4.2. Operational Overview

This section provides an informative lifecycle overview to contextualize constraint validation, invocation validation, observation, and drift handling.

1. The Intent Originator expresses an intent via the Intent Client.

2. The Intent Client submits an intent artifact to the IBS.

3. The IBS performs intent translation (Intent Transformer) to derive constraints, objectives, and candidate directives.

4. The IBS performs constraint validation (R3) in consultation with the Policy/Constraint Authority.

5. The IBS determines one or more tool/service invocations needed for fulfillment.

6. Prior to each invocation, the IBS performs invocation validation (R4), including privilege checks and parameter/constraint checks.

7. The Tool/Service Provider executes the invocation and returns results; side effects may be irreversible.

8. The Observer produces observations used by the IBS for assurance and drift detection (R6).

9. If drift or violations are detected, the IBS applies a policy-driven response (R7), such as deny, degrade, re-confirm, re-negotiate, or fallback.

5. Security Scenarios

This section describes representative security scenarios using a consistent template: Setting, Actors, Assets, Attack Sketch, Impact, and Relevant Requirements. These scenarios are not exhaustive but illustrate key threat patterns in multi-hop intent processing.

5.1. Scenario 1: Directive Tampering and Authorization Boundary Expansion

Setting:

An IBS translates an intent into derived directives (e.g., allowed rules) that traverse multiple intermediaries before reaching an execution endpoint.

Actors:

Intent Originator, Intent Client, IBS, one or more intermediaries (agents/clients), Tool/Service Provider.

Assets:

Authorization boundary, constraints/invariants, protected resources, audit evidence.

Attack Sketch:

1. An intermediary modifies derived directives to add operations or widen resource scope.

2. The modified directives are forwarded to the execution endpoint without effective detection.

3. The endpoint performs out-of-bound operations (e.g., modifying account state, accessing other parties' data, disabling safety rules).

Impact:

Privilege escalation, policy bypass, unauthorized side effects, compliance violations.

Relevant Requirements:

R1 (Provenance and Authorization Boundary Binding), R2 (Chain-of-Custody for Derived Directives), R3 (Constraint Validation), R5 (Non-Bypass Enforcement), R6 (Observability and Auditability).

6. Security Considerations

This section provides solution-agnostic security considerations mapped to the scenarios and requirements. Implementations may realize these considerations using different security mechanisms (tokens, signatures, attestation, policy engines, or protocol-level bindings).

6.1. Scenario-to-Requirement Mapping

Table 1 summarizes the primary mappings between the elaborated scenarios and security requirements. Note that these mappings are non-exhaustive; additional requirements may apply depending on deployment context.

Scenario	Primary Threats	Key Requirements
1 (Directive Tampering)	T1, T3	R1, R2, R3, R5, R6

Table 1: Scenario to Requirement Mapping

6.2. Considerations for Scenario 1 (Directive Tampering)

Scenario 1 highlights that derived directives are often more operationally powerful than the original intent text. Therefore, systems should treat derived directives as security-relevant artifacts whose integrity and authorization boundary binding should be protected across hops.

6.2.1. Overview

The core challenge is ensuring that derived directives cannot be tampered with or substituted in transit, and that execution endpoints can verify the authenticity and authorization boundary of received directives.

Binding and Custody (R1, R2): Derived directives should be bound to the intent context and authorization boundary such that unauthorized expansion or substitution is detectable or preventable across hops.

Pre-Execution Constraint Validation (R3): Even if directives appear intact, the receiver should validate that the intended actions remain within constraints and invariants before execution.

Non-Bypass Enforcement (R5): Execution endpoints should enforce checks such that direct calls cannot bypass required validation gates.

Audit Evidence (R6): Systems should produce evidence linking execution decisions to validated directives and constraints.

6.2.2. Illustrative Procedure

The following procedure is informative and solution-agnostic. Implementations may use various mechanisms (e.g., signed tokens, cryptographic binding, attestation) to achieve these objectives.

1. Directive Derivation and Binding: The IBS derives directives from an intent and associates them with the applicable authorization boundary. The IBS generates a cryptographically-protected artifact (e.g., signed token, sealed directive) that binds the directives to the intent context and authorization scope.

2. Integrity Protection for Multi-Hop Propagation: Before forwarding directives across trust boundaries, the system attaches integrity and binding evidence (e.g., digital signature, MAC, or attestation token) sufficient for downstream verification. This evidence includes the authorization boundary, constraint set, and issuer identity.

3. Reception and Verification: Upon receipt, the execution-side gate verifies the integrity and binding evidence of the received directives. This verification confirms: (a) the directives have not been tampered with or substituted, (b) the directives originate from an authorized IBS, and (c) the authorization boundary matches expected scope.

4. Constraint Re-Validation: The execution endpoint re-validates the directives against local constraints, invariants, and policy rules. This step provides defense-in-depth even if upstream validation was bypassed or compromised.

5. Enforcement and Audit: If verification or validation fails, the system denies or degrades execution under policy and records audit evidence. If successful, the system proceeds with execution and logs the binding evidence, execution decision, and outcomes for compliance assessment.

This procedure addresses T1 (Directive Tampering/Substitution) and T3 (Constraint Bypass) by establishing end-to-end integrity and validation across multi-hop processing.

6.3. General Security Considerations

Beyond the scenario-specific considerations, the following general principles apply to intent-based systems:

* Trust Boundary Awareness: systems explicitly identify trust boundaries and apply appropriate security controls at each boundary crossing.

* Defense in Depth: validation occur at multiple layers (translation, propagation, invocation, execution) to provide resilience against bypass or compromise of individual layers.

* Least Privilege: derived directives and invocations are scoped to the minimum privileges necessary for intent fulfillment.

* Fail-Safe Defaults: when validation fails or drift is detected, systems default to denying actions rather than permitting potentially unsafe operations.

* Auditability: all security-relevant decisions and events are logged with sufficient context to support post-incident investigation and compliance assessment.

7. IANA Considerations

This document has no IANA actions.

8. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC7991] Hoffman, P., "The "xml2rfc" Version 3 Vocabulary", RFC 7991, DOI 10.17487/RFC7991, December 2016, <https://www.rfc-editor.org/info/rfc7991>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

9. Informative References

[I-D.goswami-agentic-jwt] Goswami, A., "Secure Intent Protocol: JWT Compatible Agentic Identity and Workflow Management", Work in Progress, Internet-Draft, draft-goswami-agentic-jwt-00, 1 January 2026, <https://datatracker.ietf.org/doc/html/draft-goswami-agentic-jwt-00>.

[I-D.ietf-oauth-transaction-tokens] Tulshibagwale, A., Fletcher, G., and P. Kasselman, "Transaction Tokens", Work in Progress, Internet-Draft, draft-ietf-oauth-transaction-tokens-08, 2 March 2026, <https://datatracker.ietf.org/doc/html/draft-ietf-oauth-transaction-tokens-08>.

[I-D.irtf-nmrg-ibn-usecases] Yao, K., Chen, D., Jeong, J. P., Wu, Q., Yang, C., Contreras, L. M., and G. Fioccola, "Use Cases and Practices for Intent-Based Networking", Work in Progress, Internet-Draft, draft-irtf-nmrg-ibn-usecases-03, 15 March 2026, <https://datatracker.ietf.org/doc/html/draft-irtf-nmrg-ibn-usecases-03>.

[I-D.liu-oauth-a2a-profile]
Liu, P. C. and N. Yuan, "Agent-to-Agent (A2A) Profile for OAuth Transaction Tokens", Work in Progress, Internet-Draft, draft-liu-oauth-a2a-profile-00, 20 October 2025, <https://datatracker.ietf.org/doc/html/draft-liu-oauth-a2a-profile-00>.

[I-D.ni-a2a-ai-agent-security-requirements]
Yuan, N., Liu, P. C., Gao, Q., and Z. Li, "Security Requirements for AI Agents", Work in Progress, Internet-Draft, draft-ni-a2a-ai-agent-security-requirements-01, 28 February 2026, <https://datatracker.ietf.org/doc/html/draft-ni-a2a-ai-agent-security-requirements-01>.

[I-D.oauth-transaction-tokens-for-agents]
Raut, A., "Transaction Tokens For Agents", Work in Progress, Internet-Draft, draft-oauth-transaction-tokens-for-agents-04, 10 February 2026, <https://datatracker.ietf.org/doc/html/draft-oauth-transaction-tokens-for-agents-04>.

[RFC9315] Clemm, A., Ciavaglia, L., Granville, L. Z., and J. Tantsura, "Intent-Based Networking - Concepts and Definitions", RFC 9315, DOI 10.17487/RFC9315, October 2022, <https://www.rfc-editor.org/info/rfc9315>.

[RFC9316] Li, C., Havel, O., Olariu, A., Martinez-Julia, P., Nobre, J., and D. Lopez, "Intent Classification", RFC 9316, DOI 10.17487/RFC9316, October 2022, <https://www.rfc-editor.org/info/rfc9316>.

Acknowledgments

TODO

Authors' Addresses

Yuning Jiang
Huawei
Email: jiangyuning2@h-partners.com

Lun Li
Huawei
Email: lilun20@huawei.com

Yurong Song
Huawei
Email: songyurong1@huawei.com

Faye Liu
Huawei
Email: liufei19@huawei.com