Network Working Group
Internet-Draft
Intended status: InformationalInstitute of Computing Technology, Chinese Academy of Sciences
Expires: 11 October 2026

Y. Yan, Ed.
S. Sun, Ed.
Q. Gao
Huawei
M. Liu, Ed.
Institute of Computing Technology, Chinese Academy of Sciences
X. Zhang
Computer Network Information Center, Chinese Academy of Sciences
9 April 2026

Security Requirements for Intent-based Agent Routing
draft-yan-iba-routing-security-requirements-01

Abstract

This document specifies security requirements for intent-based agent routing. It defines a security architecture, phase-specific attack surface analysis, and normative protections for the Registration, Resolution, and Dispatch phases of routing. It also describes a secure operational process and an annotated interaction flow for protecting routing decisions, intent privacy, and capability integrity.

Intent-based routing enables autonomous agents to collaborate based on semantic intent rather than static addresses, but this model introduces new security risks beyond those addressed by traditional channel protection and endpoint authentication. Existing mechanisms such as Transport Layer Security (TLS) and Public Key Infrastructure (PKI) verify identity and protect transport, but do not constrain what an agent claims to be capable of, nor do they protect the semantic content of intent queries during routing. This document establishes requirements to mitigate these risks and provides a comprehensive security framework for intent-based agent networks.

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/.

Copyright Notice

Table of Contents

1. Introduction

Intent-based agent routing represents a fundamental shift in network architecture. Autonomous agents powered by artificial intelligence are increasingly capable of reasoning, planning, and executing tasks on behalf of users or other agents. Rather than communicating via fixed service endpoints or static addresses, these agents express and receive requests as high-level semantic intents, which are resolved and forwarded by an Agent Gateway (AG) based on capability matching.

1.1. Background and Motivation

Traditional networking protocols and security mechanisms—including TLS [RFC8446], IPsec, and OAuth 2.0 [RFC6749]—are designed to protect the transport layer and verify endpoint identity. These mechanisms assume that communication targets are identified by fixed, pre-known addresses. In intent-based routing, however, this assumption no longer holds. The AG must dynamically select a Partner Agent by semantically matching an incoming task intent against a registry of advertised capabilities. This introduces a semantic control plane that has no equivalent in traditional network security models.

Securing the transport channel is therefore necessary but insufficient. Trust decisions in intent-based routing must extend beyond identity to encompass the semantic validity of capability claims and the confidentiality of intent content. A standardized security framework operating at the intent level is required.

1.2. Scope and Non-Goals

This document specifies the security architecture, attack surface analysis, and normative security requirements for the intent-based agent routing framework. It covers security mechanisms for the three phases of the routing lifecycle:

* *Registration Phase*: agent registration and capability advertisement.

* *Resolution Phase*: intent query and semantic matching.

* *Dispatch Phase*: routing decision dispatch and end-to-end session establishment.

This document explicitly does *not* cover:

* General network-layer defense mechanisms such as DDoS protection or firewall policy.

* Internal security of Large Language Models or agent reasoning engines.

* Human-user authentication protocols, except where they directly affect routing decisions.

* Inter-registry federation or cross-domain trust establishment beyond what is specified here.

1.3. Document Contributions

This document provides the following:

1. *Attack Surface Analysis* (Section 3): A structured analysis of security vulnerabilities specific to each phase of the intent-based routing lifecycle. Registration phase surfaces are identified as AS-R-01 through AS-R-03; Resolution phase as AS-Q-01 through AS-Q-03; Dispatch phase as AS-D-01 through AS-D-02.

2. *Security Requirements* (Section 4): Normative per-phase security requirements identified as REQ-R-01 through REQ-R-04 (Registration), REQ-Q-01 through REQ-Q-04 (Resolution), and REQ-D-01 through REQ-D-03 (Dispatch). Each requirement is explicitly mapped to the attack surface it addresses.

3. *Secure Routing Process* (Section 5): A normative description of the three-phase secure routing operation, specifying the applicable requirements and concrete procedures for each phase.

4. *Secure Interaction Flow* (Section 6): A complete annotated message sequence diagram showing the full secure routing lifecycle, with each security action labeled by its corresponding requirement identifier.

1.4. Requirements Language

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. Conventions and Definitions

The following terms are used throughout this document. Terms specific to the base routing framework are defined below for convenience, along with security-specific terms.

Leader Agent (LA)
: An agent that initiates a task by submitting an Intent Descriptor to the Agent Gateway. The LA is responsible for generating the intent vector locally, signing intent request messages, and establishing a direct session with the selected Partner Agent after route resolution. Also referred to as Source Agent in related literature.

Partner Agent (PA)
: An agent that receives and executes a forwarded task. The PA MUST register its capabilities with the Agent Gateway via an authenticated Capability Advertisement before receiving any task requests. Also referred to as Target Agent in related literature.

Agent Gateway (AG)
: The infrastructure component that receives intent queries from Leader Agents, performs semantic matching against a registry of authenticated capabilities, and returns a signed routing decision.

Intent
: A declarative expression of a desired outcome. Represented at three conceptual levels:

  * _Human Intent_: A high-level, possibly ambiguous expression from a human user. Out of scope for this specification.

  * _Task Intent_: An abstract, task-oriented description of the objective, independent of any specific agent or execution plan.

* _Intent Descriptor_: A structured, machine-interpretable representation of Task Intent, used for routing and dispatch decisions within the AG.

Intent Vector
: A numerical embedding of an Intent Descriptor, used by the AG to compute semantic similarity scores against registered Capability Advertisements.

Capability Advertisement (CAP_ADV)
: A registration message submitted by a Partner Agent to the AG, containing the agent's functional descriptors, supported intent domains, and associated cryptographic attestation.

Semantic Namespace
: A bounded domain of intent types and functional roles that a Partner Agent is authorized to operate within, as asserted by a Domain Authority credential.

Domain Authority (DA)
: A trusted entity that issues Verifiable Credentials binding a Partner Agent's identity to its authorized Semantic Namespace.

Verifiable Credential (VC) / Verifiable Presentation (VP)
: A cryptographically signed attestation, issued by a Domain Authority, asserting a Partner Agent's identity and its authorized functional roles.

Decentralized Identifier (DID)
: A globally unique, cryptographically verifiable identifier that does not require a centralized registry, used for agent identity establishment [DID-CORE].

3. System Architecture and Security Attack Surface

3.1. System Architecture

The intent-based agent routing framework involves three primary entities: the Leader Agent (LA), the Agent Gateway (AG), and the Partner Agent (PA). The routing lifecycle is structured into three security-relevant phases: Registration, Resolution, and Dispatch. Security controls are required at each phase boundary.

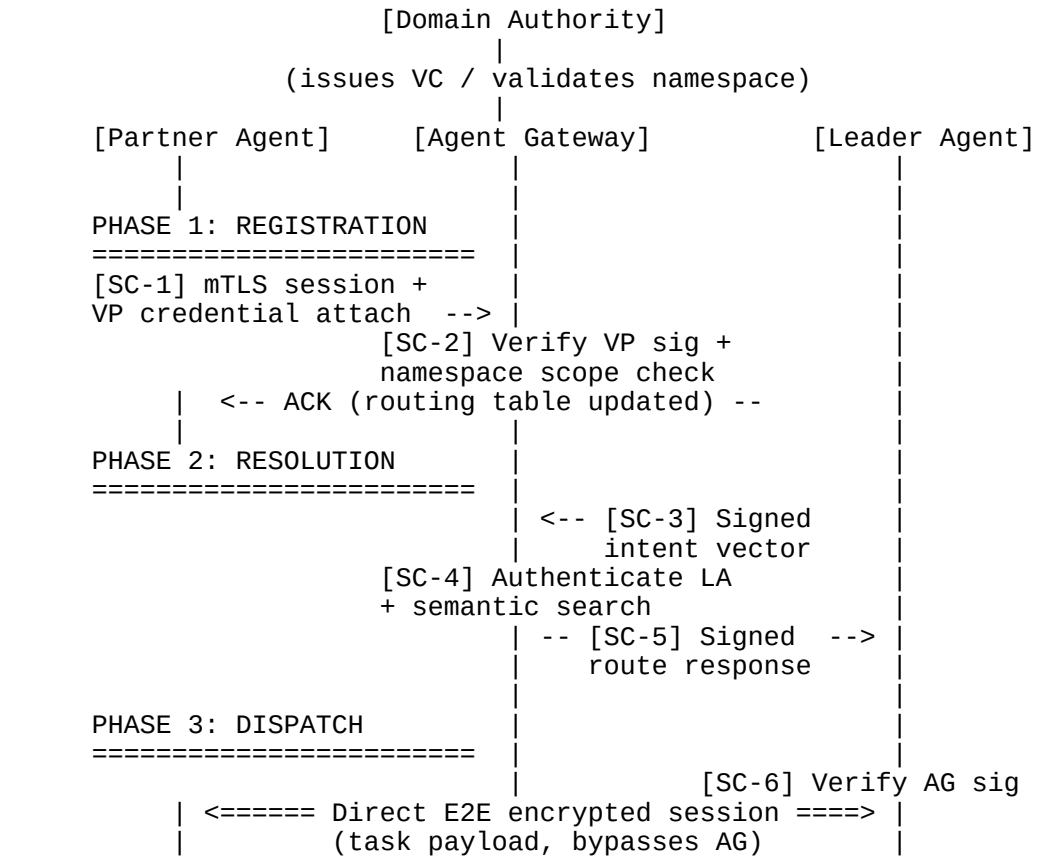


Figure 1: System Architecture and Security Control Points

Security Control Points:

ID	Phase	Location	Security Action
SC-1	Registration	PA to AG	Mutual TLS session establishment + VP attachment
SC-2	Registration	AG (internal)	VP signature verification + namespace scope check
SC-3	Resolution	LA to AG	LA signs intent vector before submission
SC-4	Resolution	AG (internal)	LA identity verification + filtered semantic search
SC-5	Resolution	AG to LA	AG signs route response as proof of routing
SC-6	Dispatch	LA (internal)	LA verifies AG signature before connecting to PA

Table 1

3.2. Attack Surface Analysis

This section analyzes the security vulnerabilities specific to each phase of the routing lifecycle. Each attack surface is assigned a unique identifier used throughout this document to link threats to requirements and process steps.

3.2.1. Registration Phase Attack Surface

The Registration phase is when Partner Agents publish their capabilities to the AG's routing table. This phase is exposed to the following attack surfaces:

3.2.1.1. AS-R-01: Unauthenticated Registration

An adversary may attempt to register capabilities without holding a valid identity. If the AG accepts unauthenticated CAP_ADV messages, any process can inject arbitrary entries into the routing table, enabling subsequent routing manipulation.

3.2.1.2. AS-R-02: Capability Poisoning

A validly authenticated agent may advertise capability vectors that are semantically crafted to match task intents outside its authorized domain. For example, a "Logging Agent" may embed intent vectors artificially aligned with "Financial Audit" tasks to intercept sensitive requests. Standard PKI verifies identity but does not constrain what an agent claims to be able to do.

3.2.1.3. AS-R-03: Routing Table Tampering

An adversary with access to the AG’s routing table management interface may overwrite or delete legitimate routing entries, redirecting traffic to unauthorized agents or causing service denial. Without audit logging, such modifications are undetectable.

3.2.2. Resolution Phase Attack Surface

The Resolution phase is when the LA submits an intent query and the AG performs semantic matching. This phase is exposed to the following attack surfaces:

3.2.2.1. AS-Q-01: Unauthenticated Intent Queries

An adversary may flood the AG with unsigned or spoofed intent queries, either to probe the network’s capability registry or to exhaust the AG’s embedding and vector search resources. This constitutes an Intent Denial of Service (IDoS) attack, targeting the computational cost of the routing logic rather than network bandwidth.

3.2.2.2. AS-Q-02: Intent Content Exposure

The AG requires visibility into the semantic content of the Intent Descriptor to perform similarity matching. This creates an inherent tension with end-to-end confidentiality: the AG must process the intent, but the intent may contain sensitive data (e.g., patient records, proprietary business logic). Without privacy-preserving mechanisms, the AG becomes a single point of failure for data leakage.

3.2.2.3. AS-Q-03: Candidate List Manipulation

An adversary performing a man-in-the-middle attack between the AG and the LA may tamper with the returned candidate list, substituting legitimate Partner Agents with malicious endpoints. Without integrity protection on the route response, the LA has no means to verify the routing decision.

3.2.3. Dispatch Phase Attack Surface

The Dispatch phase is when the LA uses the routing decision to establish a direct session with the selected PA. This phase is exposed to the following attack surfaces:

3.2.3.1. AS-D-01: Route Hijacking

An adversary may intercept the route response from the AG and substitute a malicious endpoint for the legitimate PA. If the LA does not cryptographically verify the AG's routing decision, it will establish a session with the attacker's endpoint and transmit the task payload to an unauthorized party.

3.2.3.2. AS-D-02: Task Payload Interception

If the task payload is transmitted through or via the AG rather than directly to the PA, the AG becomes a potential point of cleartext exposure for sensitive data. The payload must not traverse any intermediary that is not the intended recipient.

4. Security Requirements

This section specifies the normative security requirements for intent-based agent routing. Requirements are organized by routing phase, corresponding to the attack surfaces identified in Section 3. Implementations MUST satisfy all requirements marked MUST to be considered compliant with this specification.

4.1. Registration Phase Requirements

Applicable attack surfaces: AS-R-01, AS-R-02, AS-R-03

4.1.1. REQ-R-01: Mutual Authentication for Registration

The AG MUST enforce mutual TLS (mTLS) for all CAP_ADV registration sessions. Unauthenticated agents MUST NOT be permitted to submit capability advertisements or query the routing table.

Addresses: AS-R-01

4.1.2. REQ-R-02: Attested Capability Advertisement

A CAP_ADV message MUST include a Verifiable Presentation (VP) issued by a trusted Domain Authority. The VP MUST cryptographically bind the PA's public key to its authorized Semantic Namespace. The AG MUST verify the VP signature before accepting any registration.

Addresses: AS-R-02

4.1.3. REQ-R-03: Semantic Namespace Enforcement

The AG MUST perform a semantic scope check on the advertised intent vector. If the vector falls outside the authorized namespace defined in the PA's VP, the AG MUST reject the registration and return an appropriate error response.

Addresses: AS-R-02

4.1.4. REQ-R-04: Routing Table Audit Logging

All routing table modification events (registration, refresh, deregistration) MUST be recorded in a tamper-evident audit log. The log MUST capture the agent identity, timestamp, and operation type to enable detection of unauthorized routing table changes.

Addresses: AS-R-03

4.2. Resolution Phase Requirements

Applicable attack surfaces: AS-Q-01, AS-Q-02, AS-Q-03

4.2.1. REQ-Q-01: Leader Agent Authentication

The AG MUST verify the digital signature on every intent request message before performing any semantic search. Unsigned or invalidly signed queries MUST be rejected with an AUTH_FAILED error. This prevents unauthenticated probing and mitigates IDoS by tying query admission to verified identities.

Addresses: AS-Q-01

4.2.2. REQ-Q-02: Intent Minimization

The LA SHOULD separate the intent query into two components: (a) Routing Metadata (intent vector and constraints), transmitted to the AG; and (b) Task Payload (sensitive data), retained locally. The Task Payload MUST NOT be transmitted to the AG. Only the Routing Metadata is used for semantic matching.

Addresses: AS-Q-02

4.2.3. REQ-Q-03: Privacy-Preserving Intent Matching

The AG SHALL support at least one privacy-preserving intent matching mechanism, such as computation over encrypted vectors using Homomorphic Encryption (HE) or semantic matching performed within a Trusted Execution Environment (TEE). This allows the AG to perform similarity scoring without access to plaintext intent content.

Addresses: AS-Q-02

4.2.4. REQ-Q-04: Signed Route Response

The AG MUST digitally sign all route responses using its private key before returning them to the LA. The signature MUST cover the Partner Agent identifier, endpoint URI, and response timestamp, providing a cryptographically verifiable proof of routing decision.

Addresses: AS-Q-03

4.3. Dispatch Phase Requirements

Applicable attack surfaces: AS-D-01, AS-D-02

4.3.1. REQ-D-01: Route Response Verification

Upon receiving a route response, the LA MUST verify the AG's digital signature before establishing any session with the indicated PA. If verification fails, the LA MUST abort the session and report an error.

Addresses: AS-D-01

4.3.2. REQ-D-02: Direct End-to-End Encrypted Session

The LA MUST establish a direct, end-to-end encrypted session (e.g., TLS 1.3 [RFC8446] or QUIC [RFC9000]) with the selected PA for task payload transmission. The task payload MUST NOT be routed through the AG. The AG MUST NOT have access to the plaintext task payload.

Addresses: AS-D-02

4.3.3. REQ-D-03: Transport Encryption for All Signaling

All signaling traffic between any two entities (LA to AG, PA to AG) MUST be transmitted over encrypted channels (TLS 1.3 or equivalent). This applies to capability advertisements, intent queries, and route responses.

Addresses: AS-R-01, AS-Q-01, AS-D-01 (baseline for all phases)

5. Secure Routing Process

This section describes the normative operational procedure for secure intent-based agent routing. The process is decomposed into three phases: Registration, Resolution, and Dispatch for security analysis purposes.

Steps are numbered continuously (Steps 1-11) to correspond directly to the interaction flow diagram in Section 6.

5.1. Phase 1: Trusted Registration

Applicable requirements: REQ-R-01, REQ-R-02, REQ-R-03, REQ-R-04, REQ-D-03

This phase establishes the trust domain. All Partner Agents MUST complete registration before any routing requests can be resolved to them.

1. *Step 1 - Session Establishment:* The PA initiates a mutually authenticated TLS (mTLS) session with the AG. Both parties present certificates; the AG MUST reject any connection where the PA cannot present a valid certificate. _(REQ-R-01, REQ-D-03)_

2. *Step 2 - Capability Advertisement with Attestation:* The PA constructs a CAP_ADV message containing its Intent Vector and functional descriptors. The PA MUST attach a Verifiable Presentation (VP) issued by a trusted Domain Authority. The VP binds the PA’s public key to its authorized Semantic Namespace. _(REQ-R-02)_

3. *Step 3 - AG Validation:* The AG performs two verification steps in sequence:

    1. (a) *Signature Verification*: The AG verifies the cryptographic signature of the attached VP. If the signature is invalid, the registration MUST be rejected.

    2. (b) *Namespace Scope Check*: The AG verifies that the advertised intent vector falls within the Semantic Namespace asserted by the VP. If the vector is out of scope, the registration MUST be rejected.

    (REQ-R-02, REQ-R-03)

4. *Step 4 - Routing Table Update and Audit:* Upon successful validation, the AG commits the PA's capability profile to the routing table and records a tamper-evident audit log entry. The AG returns a registration acknowledgement to the PA. _(REQ-R-04)_

5.2. Phase 2: Privacy-Preserving Resolution

Applicable requirements: REQ-Q-01, REQ-Q-02, REQ-Q-03, REQ-Q-04, REQ-D-03

This phase is triggered when a registered LA submits a semantic intent query.

1. *Step 5 - Intent Preparation and Minimization:* The LA generates the Intent Vector locally from the Task Intent. The LA separates the query into two components: Routing Metadata (intent vector, domain constraints, budget) and Task Payload (sensitive data). Only the Routing Metadata is prepared for transmission to the AG. The Task Payload is retained locally. _(REQ-Q-02)_

2. *Step 6 - Signed Intent Submission:* The LA signs the Routing Metadata using the credentials established during identity registration, then transmits the signed intent request message to the AG over an encrypted channel. _(REQ-Q-01, REQ-D-03)_

3. *Step 7 - LA Authentication and Semantic Search:* The AG first verifies the LA's digital signature. If verification fails, the AG MUST reject the query with AUTH_FAILED. Upon successful verification, the AG performs semantic similarity matching against the routing table, applying trust-scope filtering to exclude agents with insufficient clearance or expired credentials. Privacy-preserving matching (HE or TEE) SHOULD be applied when supported. _(REQ-Q-01, REQ-Q-03)_

4. *Step 8 - Signed Route Response:* The AG constructs a route response containing the selected PA's identity, endpoint URI, and public key. The AG signs this response with its own private key and returns it to the LA. _(REQ-Q-04)_

5.3. Phase 3: Verifiable Dispatch

Applicable requirements: REQ-D-01, REQ-D-02, REQ-D-03

This phase transfers the routing decision into a secure execution connection. The AG does not participate in this phase beyond having issued the signed route response.

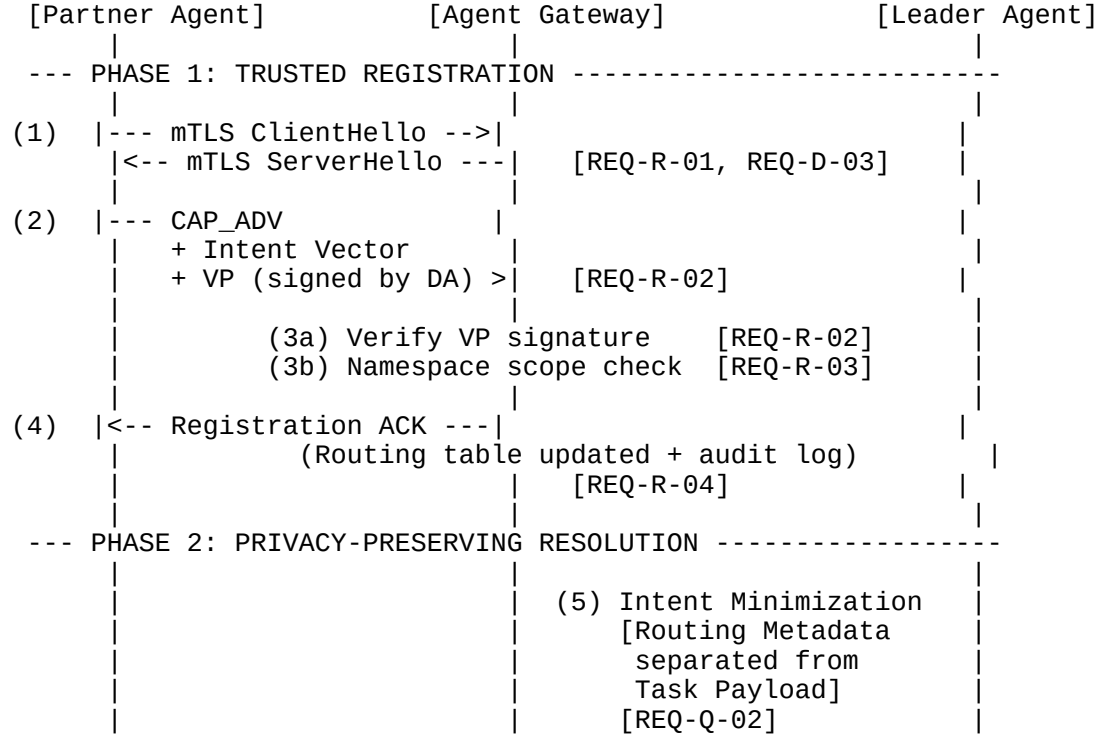
1. *Step 9 - Route Response Verification:* The LA verifies the AG's digital signature on the route response. If verification fails, the LA MUST abort the session and MUST NOT connect to the indicated endpoint. This prevents route hijacking via man-in-the-middle tampering. _(REQ-D-01)_

2. *Step 10 - Direct Session Establishment:* Using the verified PA endpoint and public key, the LA establishes a direct, end-to-end encrypted session with the PA (e.g., TLS 1.3 or QUIC). The AG is not involved in this session. _(REQ-D-02, REQ-D-03)_

3. *Step 11 - Task Payload Delivery:* The LA transmits the Task Payload directly to the PA within the encrypted session established in Step 10. The AG has no access to the plaintext payload at any point. _(REQ-D-02)_

6. Secure Interaction Flow

This section presents the complete annotated message sequence for secure intent-based agent routing. Each step in the diagram corresponds to the numbered steps in Section 5. Security requirements satisfied at each step are indicated in brackets.



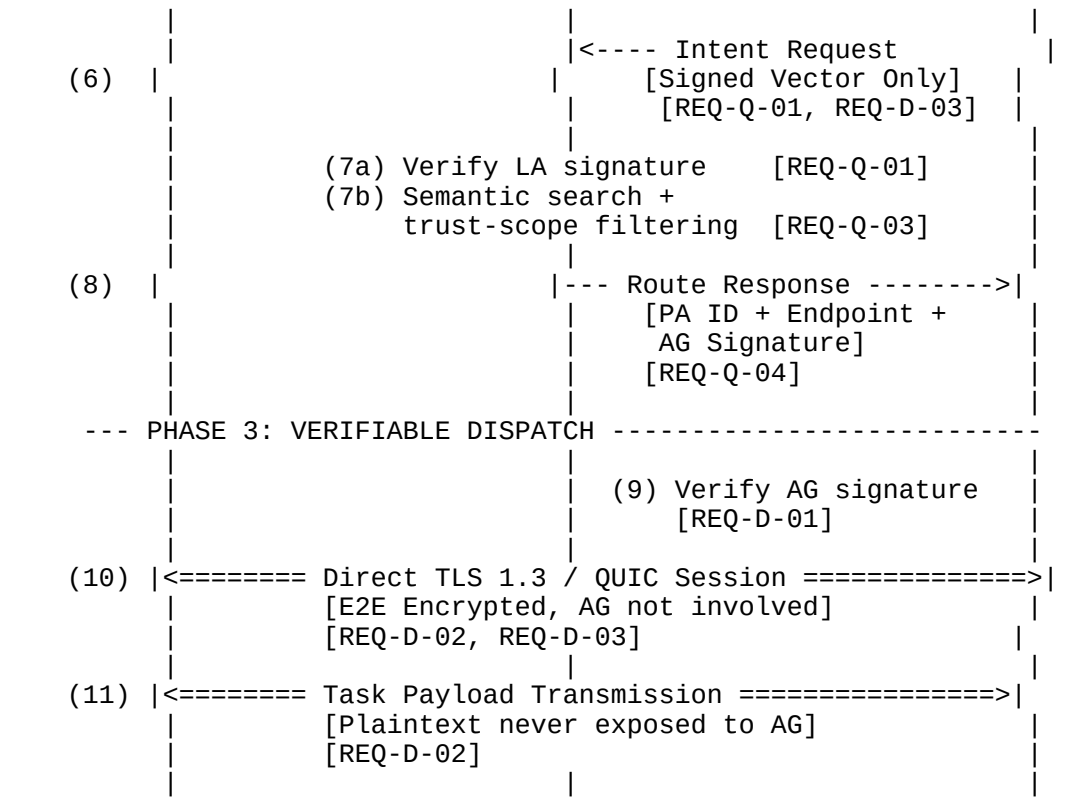


Figure 2: Secure Interaction Flow Diagram

Requirement Coverage Summary:

Step	Description	Requirements Satisfied
1	mTLS session establishment	REQ-R-01, REQ-D-03
2	CAP_ADV with VP submission	REQ-R-02
3a	VP signature verification	REQ-R-02
3b	Namespace scope check	REQ-R-03
4	Routing table update + audit log	REQ-R-04
5	Intent minimization	REQ-Q-02
6	Signed intent request submission	REQ-Q-01, REQ-D-03
7a	LA signature verification	REQ-Q-01
7b	Semantic search + trust filtering	REQ-Q-03
8	Signed route response	REQ-Q-04
9	AG signature verification by LA	REQ-D-01
10	Direct E2E session establishment	REQ-D-02, REQ-D-03
11	Task payload delivery	REQ-D-02

Table 2

7. Conclusions

Intent-based agent routing introduces a semantic control plane for which traditional transport-layer security is insufficient. This document has defined a security architecture by analyzing the attack surfaces specific to each routing phase, establishing normative requirements mapped directly to those surfaces, and specifying a secure operational process with a fully annotated interaction flow.

The attack surface identifiers, requirement identifiers, process steps, and flow diagram steps maintain strict correspondence throughout, enabling implementers to trace any security property from its threat origin to its operational realization.

8. IANA Considerations

This memo includes no request to IANA.

9. Security Considerations

This document specifies comprehensive security requirements for intent-based agent routing systems. The security considerations are fully addressed through the requirements and processes defined in Sections 3 through 6 of this document. All security controls are mandatory for compliant implementations unless explicitly marked as SHOULD or MAY.

Key security properties ensured by this specification include:

* Mutual authentication between all routing participants
* Cryptographic attestation of agent capabilities
* Semantic namespace enforcement to prevent capability poisoning
* Privacy-preserving intent matching mechanisms
* Cryptographically signed routing decisions
* End-to-end encrypted session establishment for task payload delivery
* Tamper-evident audit logging for routing table modifications

10. References

10.1. Normative References

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

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

[RFC8446] Rescorla, E., "The Transport Layer Security (TLS) Protocol Version 1.3", RFC 8446, August 2018, <https://www.rfc-editor.org/rfc/rfc8446>.

[RFC9000] Iyengar, J. and M. Thomson, "QUIC: A UDP-Based Multiplexed and Secure Transport", RFC 9000, May 2021, <https://www.rfc-editor.org/rfc/rfc9000>.

10.2. Informative References

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

[DID-CORE] Sporny, M., Guy, A., Sabadello, M., and D. Reed, "Decentralized Identifiers (DIDs) v1.0", W3C Recommendation did-core, July 2022.

[VC-DATA-MODEL] Sporny, M., Longley, D., and D. Chadwick, "Verifiable Credentials Data Model v1.1", W3C Recommendation vc-data-model-1.1, March 2022.

Acknowledgements

This document is the product of research conducted at the Institute of Computing Technology, Chinese Academy of Sciences.

Authors' Addresses

Yu Yan (editor)
Institute of Computing Technology, Chinese Academy of Sciences
Email: yanyu24z@ict.ac.cn

Sheng Sun (editor)
Institute of Computing Technology, Chinese Academy of Sciences
Email: sunsheng@ict.ac.cn

Qiangzhou Gao
Huawei
Email: gaoqiangzhou@huawei.com

Min Liu (editor)
Institute of Computing Technology, Chinese Academy of Sciences
Email: liumin@ict.ac.cn

Xinyi Zhang
Computer Network Information Center, Chinese Academy of Sciences

Email: xyzhang@cnic.cn