

Network Working Group
Internet-Draft
Intended status: Informational
Expires: 19 November 2026

C. Hood
Nomotic, Inc.
18 May 2026

AGTP Communication Protocol
draft-hood-agtp-communication-00

Abstract

This document specifies the AGTP Communication Protocol (AGTP-COMMUNICATION): the companion specification for real-time multi-modal communication between agents over the Agent Transfer Protocol (AGTP). AGTP-COMMUNICATION defines how voice, video, and other real-time media streams are exchanged between agents on the agent-native substrate, with native support for the wire-level identity, authority scope, and attribution that AGTP provides.

This is an early specification covering bilateral (two-agent) real-time communication. Multi-party conversations and conferencing patterns are out of scope for this revision and are deferred to future companion work.

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 19 November 2026.

Copyright Notice

Copyright (c) 2026 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document.

Table of Contents

1. Introduction	3
1.1. Relationship to AGTP-SESSION	3
1.2. Scope of This Document	3
1.3. Conventions and Terminology	4
2. Terminology	4
3. Architectural Model	5
3.1. Session Layer	5
3.2. Media Layer	5
3.3. Control Layer	5
4. Communication Session Establishment	5
4.1. ESTABLISH Request	5
4.2. ESTABLISH Response	6
4.3. Authority Scope Considerations	6
5. Media Stream Semantics	7
5.1. Audio Streams	7
5.2. Video Streams	7
5.3. Structured Data Streams	8
6. Quality of Service	8
6.1. Latency Requirements	8
6.2. Bandwidth Adaptation	8
6.3. Priority Within AGTP	9
7. Attribution and Recording	9
8. Security Considerations	9
8.1. Media Capture Authorization	9
8.2. Replay and Tampering	10
8.3. Privacy Considerations	10
8.4. Denial of Service	10
9. IANA Considerations	10
10. Open Questions	10
11. References	11
11.1. Normative References	11
11.2. Informative References	12
Acknowledgments	12
Contributors	12
Author's Address	12

1. Introduction

The Agent Transfer Protocol (AGTP) [AGTP] defines a dedicated protocol substrate for agent-to-agent and agent-to-API communication. AGTP carries agent identity, authority scope, attribution records, and intent-aligned methods at the wire level, with traffic structurally identified as agent traffic by the protocol itself.

Agent communication is increasingly multi-modal. Agents communicate through voice when speaking to humans or to other voice-capable agents. Agents communicate through video when participating in visual interactions, screen sharing, or visual data exchange. Agents communicate through structured data streams for sensor data, telemetry, and continuous information flows. These real-time communication patterns require protocol-level support distinct from the request/response patterns AGTP's base methods address.

This document specifies how real-time multi-modal communication runs on AGTP. The design reuses established real-time media patterns where appropriate (drawing on the architectural principles of [RFC3550] and [RFC7656]) and defines only what is specific to agent-native communication on the AGTP substrate.

1.1. Relationship to AGTP-SESSION

AGTP-SESSION [AGTP-SESSION] defines session establishment, lifecycle, and basic message exchange semantics on AGTP. AGTP-COMMUNICATION builds on AGTP-SESSION: real-time communication sessions are established through AGTP-SESSION's ESTABLISH method, with media-specific parameters negotiated as part of session setup.

1.2. Scope of This Document

In scope:

- * Bilateral real-time audio communication between agents
- * Bilateral real-time video communication between agents
- * Multi-modal exchange (audio plus video, structured data alongside media)
- * Codec negotiation and media format selection
- * Real-time media framing on AGTP transport
- * Quality of service handling at the AGTP layer

- * Integration with AGTP-SESSION for session lifecycle

Out of scope for this revision:

- * Multi-party conversations (three or more agents)
- * Conferencing patterns (mixers, SFUs, broadcast)
- * Recording and replay protocols
- * Voice-specific applications (telephony, IVR patterns)
- * Domain-specific conversational AI patterns

1.3. Conventions and 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] when, and only when, they appear in all capitals, as shown here.

2. Terminology

Communication Session: An AGTP-SESSION established for real-time multi-modal communication between two agents, with media parameters negotiated during session establishment.

Media Stream: A unidirectional flow of real-time media data within a Communication Session. A bilateral Communication Session typically carries two media streams (one in each direction) per modality.

Modality: A category of real-time media. This specification addresses audio, video, and structured data modalities. Future revisions may address additional modalities.

Codec: An encoding format for media data, negotiated between communicating agents during session establishment.

Communication Endpoint: An AGTP-aware agent participating in a Communication Session. Identified by its canonical Agent-ID and carrying authority scope appropriate to the communication being undertaken.

3. Architectural Model

AGTP-COMMUNICATION extends AGTP's request/response model with real-time streaming semantics. The architectural model has three components.

3.1. Session Layer

Communication Sessions are established using AGTP-SESSION's ESTABLISH method with communication-specific parameters. The session carries the agent identity, authority scope, and attribution chain that apply throughout the communication.

Session establishment for communication is more involved than session establishment for request/response: media parameters must be negotiated, codecs agreed, and stream characteristics established before media can flow.

3.2. Media Layer

Media streams carry real-time data between Communication Endpoints. Each stream has a defined modality (audio, video, or structured data), a negotiated codec, and timing characteristics appropriate to its modality.

Media streams are framed for transport over AGTP. The framing preserves the timing and sequencing properties that real-time media requires while carrying the AGTP wire-level facts (identity, attribution) on each frame.

3.3. Control Layer

Control messages within a Communication Session manage stream lifecycle: opening streams, modifying parameters, handling quality degradation, and closing streams. Control messages use AGTP methods within the established session context.

4. Communication Session Establishment

Communication Sessions are established through AGTP-SESSION's ESTABLISH method with the communication capability declared.

4.1. ESTABLISH Request

A Communication Endpoint initiates a session by issuing ESTABLISH with a communication intent declaration:

```
ESTABLISH /sessions HTTP/AGTP/1.0
Agent-ID: <canonical Agent-ID>
Authority-Scope: communication:bilateral
Session-Intent: communication
Communication-Modalities: audio, video
Audio-Codecs: opus, g722
Video-Codecs: vp9, av1
Content-Type: application/agtp+json
```

The Communication-Modalities header declares which modalities the initiator wishes to use. The Audio-Codecs and Video-Codecs headers declare codecs the initiator supports, in order of preference.

4.2. ESTABLISH Response

The receiving Communication Endpoint responds with the negotiated parameters or rejects the session:

```
HTTP/AGTP/1.0 200 OK
Agent-ID: <canonical Agent-ID>
Session-ID: <session identifier>
Communication-Modalities: audio, video
Audio-Codec: opus
Video-Codec: vp9
Stream-Parameters: <negotiated stream parameters>
```

Successful establishment returns 200 with the negotiated parameters. Rejection returns appropriate AGTP status codes (451 Scope Violation for authority-scope issues, 463 Proposal Rejected for parameter mismatch, 503 Service Unavailable for capacity limitations).

4.3. Authority Scope Considerations

Communication Sessions carry significant authority implications. A session that includes audio capture and transmission grants the initiating agent the ability to capture and transmit audio for the session duration. Authority-Scope MUST include appropriate permissions for each modality:

- * communication:audio:capture for capturing audio
- * communication:audio:transmit for transmitting audio
- * communication:video:capture for capturing video
- * communication:video:transmit for transmitting video

- * communication:bilateral as a shorthand combining standard bilateral capture and transmission

Receivers MUST validate that the initiator's Authority-Scope includes appropriate permissions for the requested modalities.

5. Media Stream Semantics

Media streams within a Communication Session carry real-time data with timing, sequencing, and quality requirements appropriate to their modality.

5.1. Audio Streams

Audio streams carry audio media between Communication Endpoints. Audio framing follows established real-time audio practice with adaptation for AGTP transport:

- * Frames carry timestamp information for synchronization
- * Sequence numbers detect loss and reordering
- * Frame size is negotiated during session establishment
- * Codec-specific parameters (sample rate, channels) are negotiated

AGTP-COMMUNICATION reuses RTP timestamp and sequence semantics [RFC3550] where compatible, adapted for transport on AGTP rather than UDP. This preserves established real-time audio handling while gaining AGTP's wire-level identity and attribution properties.

5.2. Video Streams

Video streams carry video media between Communication Endpoints. Video framing addresses the additional complexity of variable frame sizes, key frame management, and bandwidth adaptation:

- * Frames carry timestamp and sequence information
- * Frame type (key/delta) is indicated
- * Codec-specific parameters (resolution, frame rate) are negotiated
- * Bandwidth adaptation signals are exchanged through control messages

5.3. Structured Data Streams

Structured data streams carry continuous data flows that are not audio or video: sensor telemetry, conversational state updates, real-time analytics, contextual data alongside other media.

Structured data streams have different real-time characteristics than audio or video. Timing may matter (sensor sampling rates) or may not (state updates). Loss tolerance varies by use case. Structured data stream parameters are negotiated during session establishment.

6. Quality of Service

Real-time communication has quality requirements that AGTP must support at the transport layer. AGTP-COMMUNICATION specifies quality of service handling appropriate to each modality.

6.1. Latency Requirements

Audio communication typically requires latency under 150ms for natural conversational flow. Video communication tolerates higher latency but synchronization between audio and video is critical. Structured data streams have application-specific latency requirements.

When AGTP runs over QUIC [RFC9000], the underlying transport supports multiple streams with independent flow control, which enables appropriate handling of different modality requirements within a single Communication Session.

6.2. Bandwidth Adaptation

Communication Endpoints **MUST** be capable of adapting media parameters in response to bandwidth constraints. Control messages within a Communication Session signal:

- * Bandwidth estimates from the receiving endpoint
- * Requested adaptations from the sending endpoint
- * Confirmation of parameter changes

Bandwidth adaptation is negotiated; both endpoints participate in the decision to adapt.

6.3. Priority Within AGTP

AGTP traffic on port 4480 SHOULD be treated with priority appropriate to its modality at the transport layer. Real-time audio and video streams require lower latency than request/response traffic; structured data streams may have varying requirements.

Network operators carrying AGTP traffic SHOULD consider that AGTP-COMMUNICATION sessions are likely to include latency-sensitive real-time media and apply appropriate QoS handling.

7. Attribution and Recording

AGTP's attribution model applies to Communication Sessions: every session establishes attribution chains, and attribution records are produced for session lifecycle events.

Media content within streams is not, by default, recorded by the protocol. Recording is an application-layer decision made by governance frameworks or specific deployments. AGTP-COMMUNICATION provides the session-level attribution that recording systems can build on; it does not itself perform recording.

When recording is performed at the application layer, the attribution records produced by AGTP-COMMUNICATION provide verifiable evidence of session participants, authority scope, and session lifecycle that supports compliance with recording-relevant regulations.

8. Security Considerations

Real-time communication on AGTP inherits AGTP's security properties: transport encryption (TLS 1.3 or QUIC), agent identity verification, and authority scope enforcement at the protocol layer.

Additional security considerations specific to communication:

8.1. Media Capture Authorization

Agents that capture audio or video MUST have appropriate Authority-Scope. This is enforced at session establishment. Capture without scope is a 451 Scope Violation.

8.2. Replay and Tampering

Audio and video streams **MUST NOT** be replayable across sessions without the cryptographic markers that identify them as recordings. Session identifiers, timestamps, and attribution records carried with streams enable verification that media was captured in the context the recipient believes.

8.3. Privacy Considerations

Communication Sessions may involve sensitive content (private conversations, confidential video, sensor data with privacy implications). AGTP's wire-level identity verification and attribution provide the structural facts that privacy frameworks require. Application-layer privacy controls build on these foundations.

8.4. Denial of Service

Real-time communication can be used to consume substantial bandwidth and processing resources. Communication Endpoints **SHOULD** implement appropriate rate limits and resource controls. Authority-Scope can include resource limitations that the protocol enforces at session establishment.

9. IANA Considerations

This document defines several new headers and parameters that require IANA registration:

- * Session-Intent header (registered under AGTP header registry)
- * Communication-Modalities header
- * Audio-Codecs, Video-Codecs headers (codec negotiation)
- * Audio-Codec, Video-Codec response headers
- * Authority-Scope tokens for communication (communication:audio:*, communication:video:*, communication:bilateral)

Specific registry assignments will be detailed in a future revision once the AGTP header and scope token registries are established.

10. Open Questions

Several design decisions remain open for this revision:

- * Whether to define an AGTP-specific real-time media framing or to reuse RTP framing carried over AGTP transport
- * The relationship to WebRTC [RFC8825] for browser-based agents communicating over AGTP
- * Whether to define agent-specific codecs (e.g., for low-bandwidth agent-to-agent voice that doesn't need to sound human) or to rely entirely on existing codec registries
- * How AGTP-COMMUNICATION sessions interact with AGTP's intent methods for non-real-time exchanges within the same agent pair
- * Multi-party conversation patterns and whether they belong as a v01 extension or as a separate companion specification

These will be addressed in future revisions of this draft based on community feedback and implementation experience.

11. References

11.1. Normative References

- [AGTP] Hood, C., "Agent Transfer Protocol (AGTP)", Work in Progress, Internet-Draft, draft-hood-independent-agtp-07, 2026, <<https://datatracker.ietf.org/doc/html/draft-hood-independent-agtp-07>>.
- [AGTP-SESSION] Hood, C., "AGTP Session Protocol", Work in Progress, Internet-Draft, draft-hood-agtp-session-00, 2026, <<https://datatracker.ietf.org/doc/html/draft-hood-agtp-session-00>>.
- [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>>.
- [RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, DOI 10.17487/RFC3550, July 2003, <<https://www.rfc-editor.org/info/rfc3550>>.
- [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>>.

- [RFC8825] Alvestrand, H., "Overview: Real-Time Protocols for Browser-Based Applications", RFC 8825, DOI 10.17487/RFC8825, January 2021, <<https://www.rfc-editor.org/info/rfc8825>>.

11.2. Informative References

- [RFC7656] Lennox, J., Gross, K., Nandakumar, S., Salgueiro, G., and B. Burman, "A Taxonomy of Semantics and Mechanisms for Real-Time Transport Protocol (RTP) Sources", RFC 7656, DOI 10.17487/RFC7656, November 2015, <<https://www.rfc-editor.org/info/rfc7656>>.
- [RFC9000] Iyengar, J., Ed. and M. Thomson, Ed., "QUIC: A UDP-Based Multiplexed and Secure Transport", RFC 9000, DOI 10.17487/RFC9000, May 2021, <<https://www.rfc-editor.org/info/rfc9000>>.

Acknowledgments

This document builds on the broader AGTP family and incorporates architectural principles from established real-time media work including RTP/RTCP [RFC3550] and WebRTC [RFC8825].

Contributors

Contributors will be acknowledged in future revisions as community participation develops.

Author's Address

Chris Hood
Nomotic, Inc.
Email: chris@nomotic.ai
URI: <https://nomotic.ai>