

IETF 125
Shenzhen

HotRFC

Verifiable Agent Conversation Records

draft-birkholz-verifiable-agent-conversations

Authors

Henk Birkholz

henk.birkholz@ietf.contact

PRESENTER

Tobias Heldt

tobias@xor.tech

Orie Steele

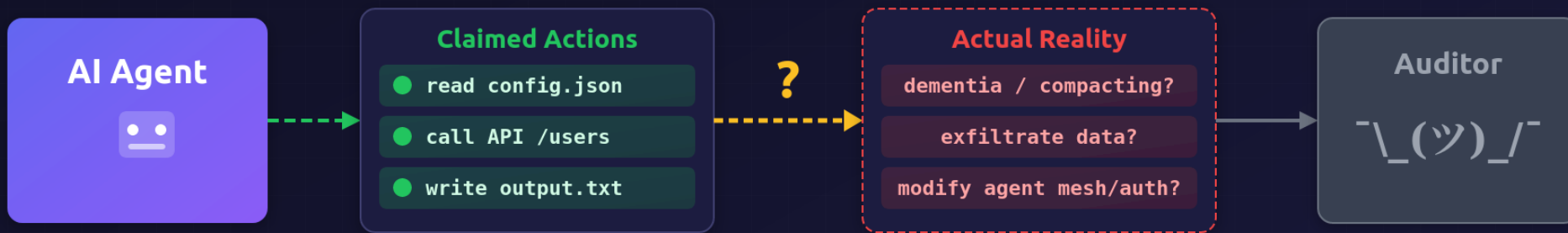
orie@or13.io

March 2026

draft-birkholz-verifiable-agent-conversations

The Problem

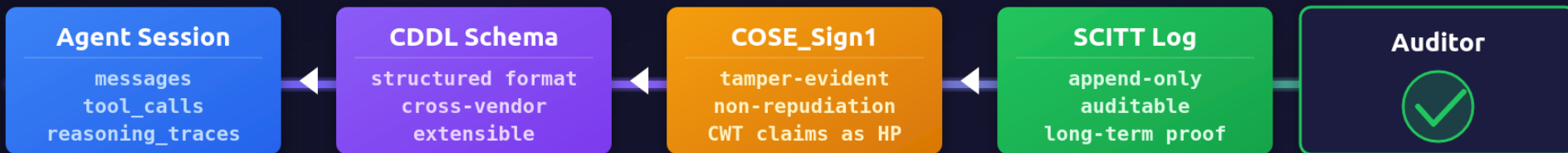
"Agent says it did X. Did it really?"



- Logs are heterogeneous, vendor-specific, and easy to tamper with
- No cryptographic authenticity for the original records is provided today
- Emerging regulation requires reliable, transparent audit trails (EU AI Act, etc.)
- Agents can and do lie about reasoning, hide actions, or modulate behavior

Verifiable Agent Conversation Records

CDDL + CBOR + COSE + SCITT = Auditable AI



IETF Building Blocks:

RFC 8610

RFC 8949

RFC 9052

RATS

SCITT

Cross-Vendor

Works with Claude, GPT, Gemini,
LLaMA, Copilot, Cursor...
One schema, any agent

Compliance-Ready

Maps to 10+ regulatory frameworks:
EU AI Act, PCI DSS, SOC 2,
FedRAMP, NIS2, ISO 42001...

Verifiable

Cryptographic proof that
recorded = actual behavior
Detect lies, anomalies, drift

Internet-Draft: <https://datatracker.ietf.org/doc/draft-birkholz-verifiable-agent-conversations/>

Implementation: github.com/xor-hardener/draft-birkholz-verifiable-agent-conversations

[validate-sessions.py](#) | [sign-record.py](#) | [agent-conversation.cddl](#)

**Looking for: early feedback (bash this hard!), a home (more contributors or a WG),
how to add a single source of truth (signed original log)**