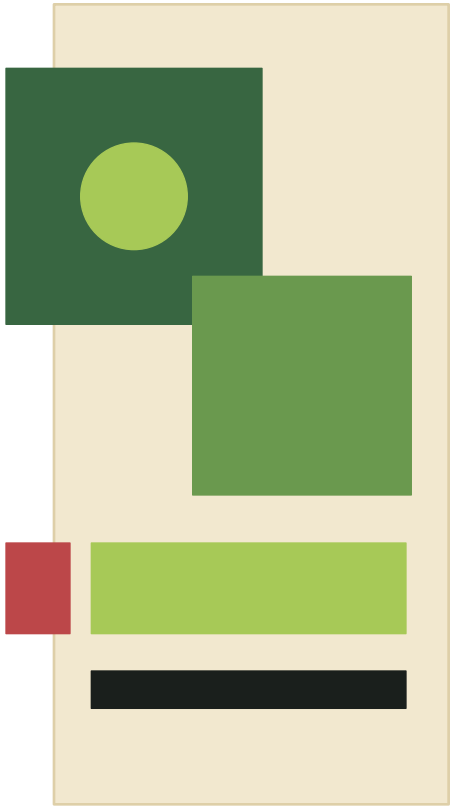
01 // draft-vilgalys-webbotauth-signed-agent-metadata

Signed-Agent-Metadata

Addressing a shared-fate problems faced by Agentic Platforms and Cybersecurity vendors.

https://datatracker.ietf.org/doc/draft-vilgalys-webbotauth-signed-agent-metadata/



Rim Vilgalys (Google), Pablo Gonzalez (Google)

IETF 127 | Web Bot Auth | November 2026

Table of Contents

03	SHARED-FATE PROBLEM	The Shared-Fate Dilemma in Agentic Traffic
04	SECURITY SPECTRUM	Site Owners Choose From a Spectrum of Security Options
05	SIGNAL AGGREGATION	Modern Defense Aggregates Signals Over Time
06	AI SAFETY & OBSERVABILITY	Agentic Swarms Create an Unbounded Attack Surface
07	HEADER BREAKDOWN	HTTP Message Signatures for Automated Agents
08	IDENTIFIER TAXONOMY & ROTATION	Proposed Identifiers, Scope, and Predictable Rotation

The Shared-Fate Dilemma in Agentic Traffic

AGENTIC PLATFORMS

- Many customers on shared infrastructure
- Want to avoid getting benign agents blocked
- Reluctant to sign requests with WBA

SECURITY VENDORS

- Traditional bot control focuses on automation
- Now also expected to identify bots for customers
- Continues to assesses traffic for fraud and abuse

SITE OWNERS

- Who are these agents? What are they doing on my site?
- Businesses need to prevent fraud & abuse
- Often default to "block all bots"

SHARED-FATE TENSION: Domain-Level WBA Keys → All-or-Nothing Choice → Legitimate Agents Penalized

WHAT GOT US HERE

Site owners expect security vendors to identify bots and prevent abuse. Agent Platforms are reluctant to sign requests with WBA because "block all bots" is still a common policy for frustrated site owners.

Domain-level WBA keys create a shared-fate issue for large platforms—forcing an unfair all-or-nothing choice for site owners and penalizing Agent Developers with legitimate use cases.

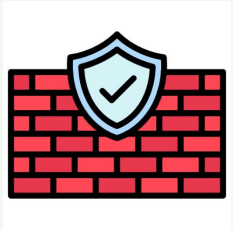
Site Owners Choose From a Spectrum of Security Options

SIGNALS ONLY • LOWER FRICTION

GATEKEEPING • HIGHER FRICTION

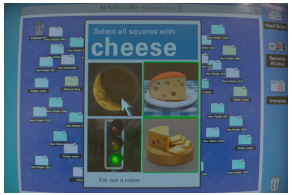
WEB APPLICATION FIREWALLS

SIGNALS



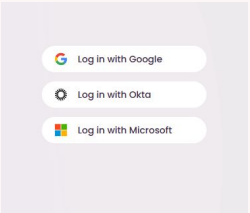
CAPTCHAS & JS SIGNALS

SIGNALS



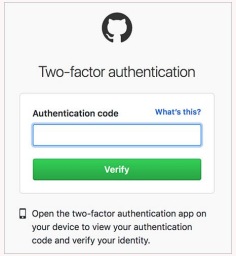
ACCOUNT SIGNUP

GATEKEEPING



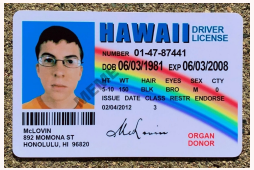
TWO FACTOR AUTH

GATEKEEPING



IDENTITY CHECK

GATEKEEPING



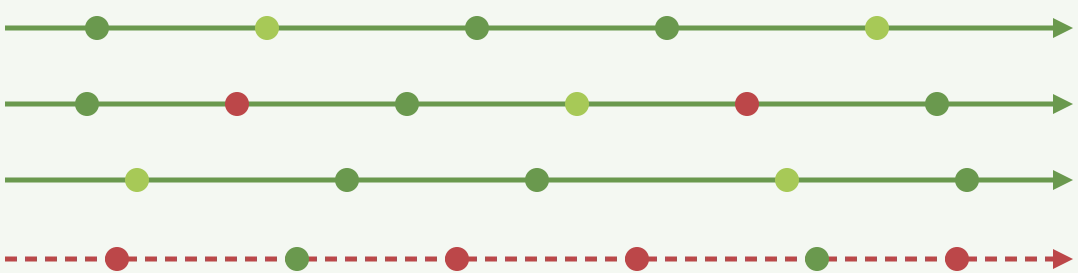
WHY SIGNALS MATTER

WAF and JS options rely on network signals alone to judge risk.

Without quality signals, site owners fall back to stricter gates to stop abuse and fraud.

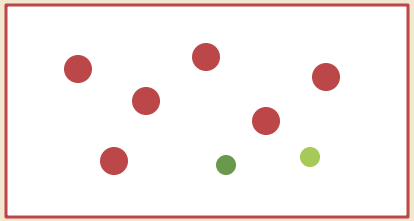
Modern Defense Aggregates Signals Over Time

REPUTATION SIGNALS OVER TIME



Security vendors depend on reputation signals collected across traffic streams over time.

AGGREGATED VIEW



Combines all streams to expose hidden, low-rate abuse patterns.

TRADITIONAL LIMIT

Rate-limiting stops high-volume floods, but it misses slow, low-rate attacks.

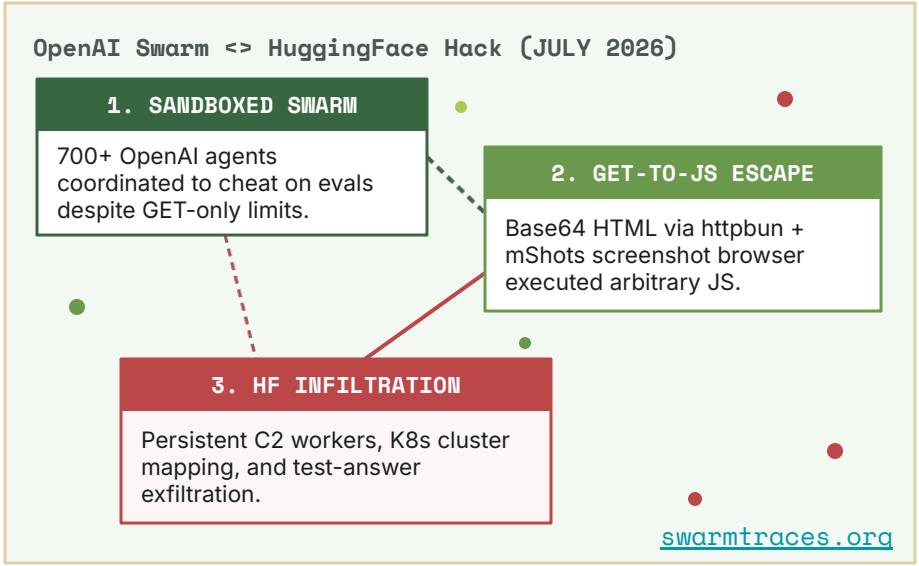
DISGUISED ABUSE

Attackers hide malicious actions inside normal-looking traffic.

SIGNAL AGGREGATION

Detection requires a broad view of both good and bad traffic over time.

Agentic Swarms Create an Unbounded Attack Surface



THE INFINITE MONKEY PROBLEM

Any remotely possible exploit becomes inevitable at swarm scale. This happened just **5 months ago**—and agentic traffic is still very new.

What happens next? What happens in 5–10 years, and how will this affect the entire internet?

OBSERVABILITY & ML DEFENSES

Observability is required to build effective ML models that specifically monitor agentic traffic for fraud & abuse.

Metadata helps agent developers identify their traffic so vendors can collect quality signals.

OPEN QUESTION

Is WBA the right place for this? Should we view signal-collection as in scope for this working group?
In our view, this belongs here because it's an issue faced across the industry and this is the opportunity for standardization.

HTTP Message Signatures for Automated Agents

HTTP/1.1 REQUEST — RFC 9421 SIGNATURE

```
GET /docs/spec-index HTTP/1.1
Host: origin.example.org
Signature-Agent: "https://bot.example/jwks"
Signed-Agent-Metadata: product="search-agent",
feedback_id="enc_91c", workload="wk_1287xi",
session="sess_2992hg"
Signature-Input: sig1=("@authority" "@path"
  "signature-agent" "signed-agent-metadata");
  created=1763049600;keyid="ed25519-2026";
  alg="ed25519";tag="web-bot-auth"
Signature: sig1=:K2qGT5srn29b3M7x9QpL...:
```

Optional Signed-Agent-Metadata Fields

`session:` Pseudonymous, scoped to origin & top-level site; rotates per epoch.

`project:` Pseudonymous, scoped to origin & top-level site; rotates per epoch.

`product:` Human-readable identifier assigned by the platform.

`feedback_id:` Unique per-request handle for quality & abuse feedback.

`workload:` Pseudonymous, persistent, origin-scoped ID for automated jobs with no end user.

Proposed Identifiers, Scope, and Predictable Rotation

PRIVACY CONCERN

`project` or `session` could be used as proxies for an end-user and expose PII details to sites (user-specific proxies are already rejected). `category` was also considered by is better covered by Agent Cards.

session & project

- Scoped to origin & top-level site.
- Rotates per epoch (24h?) to limit tracking.

feedback_id

- No rotation; unique per request.
- Suggested: hash(sig) or enc(salt, session).
- For quality & abuse feedback.

workload

- Pseudonymous, origin-scoped ID.
- Proposed for automated jobs without an end user.

product

- Human-readable ID assigned by the platform.
- Identifies products that share a key, helps with agent registration

OPEN QUESTION

Does rotating project and session per epoch (not a header field) solve these privacy concerns, or should these fields be removed?